

US-60-191-637-4899

Query Match 42.0%; Score 50; DB 24; Length 1252;  
Best Local Similarity 38.1%; Pred. No. 1.5e+02;  
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 ANNHVIVEAMSDNDPYLHD 21  
| : : | | | : : |  
DB 381 ARNYKRVAEWMDNYKRYVD 401

## RESULT 15

PCT-US01-01309-222  
; Sequence 222, Application PC/TUS0101309  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc., et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P1205PCT  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 384  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 222  
; LENGTH: 401  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (37)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (60)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (62)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (66)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (69)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (71)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (73)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (76)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (177)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
PCT-US01-01309-222

Query Match 40.8%; Score 48.5; DB 1; Length 401;  
Best Local Similarity 47.4%; Pred. No. 70;  
Matches 9; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

QY 5 VSTIVEAMSDN--DPPYLH 20  
| : : | | | : : |  
DB 46 VALLEYWKDNRFDDEXPLH 64

Search completed: March 27, 2002, 14:20:26  
Job time: 1573 sec









OY 1 AINHSLIAMSNDPOY 18  
| | | | | | | : | | | |  
Db 665 ankhlstledwngkdpqy 682

RESULT 4

AG42078  
ID AG42078 standard; Protein: 473 AA.

AC AAG42078;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52432.

KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PE 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 07-MAY-1999; 99US-0132486.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 19-MAY-1999; 99US-0134768.

PR 20-MAY-1999; 99US-0134941.

PR 21-MAY-1999; 99US-0135124.

PR 24-MAY-1999; 99US-0135353.

PR 25-MAY-1999; 99US-0135629.

PR 27-MAY-1999; 99US-0136021.

PR 28-MAY-1999; 99US-0136382.

PR 01-JUN-1999; 99US-0136782.

PR 03-JUN-1999; 99US-0137222.

PR 04-JUN-1999; 99US-0137528.

PR 07-JUN-1999; 99US-0137502.

PR 08-JUN-1999; 99US-0137724.

PR 10-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 14-JUN-1999; 99US-0138847.

PR 16-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 21-JUN-1999; 99US-0139763.

PR 22-JUN-1999; 99US-0139817.

PR 23-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144684.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145066.

PR 21-JUL-1999; 99US-0145086.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145916.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0148426.  
PR 20-AUG-1999; 99US-0148722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161356.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161933.  
PR 29-OCT-1999; 99US-0162142.

Query Match 43.1%; Score 50; DB 21; Length 473;  
Best Local Similarity 47.8%; Pred No. 10;

Matches 11; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

QY 1 AINHSLEAWSDN-DDPOYKND 21  
DB 160 algnavllstswseenygdpsendk 182

RESULT 5

AAG42077  
ID AAG42077 standard; Protein; 474 AA.  
XX  
AC AAG42077;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52431.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PF 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 43.1%; Score 50; DB 21; Length 474;  
Best Local Similarity 47.8%; Pred. No. 10;  
Matches 11; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

Oy 1 AINHSLIFAMSDN--DPOYKMD 21  
||| :||:| |||  
Db 161 aigmavliswseynygdpsend 183

RESULT 6  
AA642076  
ID AA642076 standard; Protein: 522 AA.

XX AA642076;  
AC  
XX  
DT 18-OCT-2000 (first entry)  
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52430.  
XX  
KW Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PF 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123160.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148345.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.

PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157733.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	43.1%	Score 50	DB 21	Length 522
Best Local Similarity	47.8%	Pred. No. 12		
Matches 11	Conservative 3	Mismatches 7	Indels 2	Gaps 1
QY	1 AINHLSTLEAWSDN--DPQYKND 21			
	11 : : : : : 11 11 11			
Db	209 a1gmav11swsenygdpsenkD 231			
RESULT 7				
ID	AAB93916			
XX	AAB93916 standard; Protein: 418 AA.			
XX	AAB93916;			
XX	26-JUN-2001 (first entry)			
DT				
XX	Human protein sequence SEQ ID NO:13888.			
DE				
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy.			
KW				
XX	Homo sapiens.			
OS				
XX				
PN	Ep1074617-A2.			

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602  
full-length cDNAs defined in the specification, and for the detection  
and/or diagnosis of the abnormality of the proteins encoded by the  
full-length cDNAs -

Claim 8; SEQ ID 13888; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602  
full-length cDNAs defined in the specification. Where a primer set  
comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
to the complementary strand of a polynucleotide which comprises one of  
the 5602 nucleotide sequences defined in the specification, where the  
oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
of an oligonucleotide comprising a sequence complementary to the  
complementary strand of a polynucleotide which comprises a 5'-end  
sequence and an oligonucleotide comprising a sequence complementary to a  
polynucleotide which comprises a 3'-end sequence, where the  
oligonucleotide comprises at least 15 nucleotides and the combination of  
the 5'-end sequence/3'-end sequence is selected from those defined in  
the specification. The primer sets can be used in antisense therapy and  
in gene therapy. The primers are useful for synthesizing polynucleotides,  
particularly full-length cDNAs. The primers are also useful for the  
detection and/or diagnosis of the abnormality of the proteins encoded by  
the full-length cDNAs. The primers allow obtaining of the full-length  
cDNAs easily without any specialized methods. AAH03166 to AAH13628 and  
AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
AAB95983 represent human amino acid sequences; and AAH13629 to AAH13632  
represent oligonucleotides, all of which are used in the exemplification  
of the present invention.

Sequence 418 AA;

Query Match	Similarity	42.2%	Score	49	DB	22	Length	418	
Best Local	Similarity	52.9%	Pred.	NO.	13				
Matches	9	Conservative	3	Mismatches	5	Indels	0	Gaps	0
QY	5	LSILEAWSDNDPQYNKD	21						
	:	:	:						
Db	107	lmvlagasefdpqynkd	123						
RESULT	8								
AAB92987									
ID	AAB92987	standard; Protein; 531 AA.							
XX									
XX	AAB92987;								
XX									
DT	26-JUN-2001	(first entry)							
XX									
DE	Human protein sequence SEQ ID NO:11710.								
XX									
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.								
XX									
OS	Homo sapiens.								

XX	EPI074617-A2.
PM	
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	
PA	(HELI-) HELIX RES INST.
XX	
P1	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
P1	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	
WPI:	2001-318749/34.
PT	
PT	Primer sets for synthesizing polynucleotides, particularly the 5602
PT	full-length cDNAs defined in the specification, and for the detection
PT	and/or diagnosis of the abnormality of the proteins encoded by the
PT	full-length cDNAs -
PS	
XX	Claim 8; SEQ ID 117A10; 2537P + CD ROW; English.
CC	
CC	The present invention describes primer sets for synthesising 5602
CC	full-length cDNAs defined in the specification. Where a primer set
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC	to the complementary strand of a polynucleotide which comprises one of
CC	the 5602 nucleotide sequences defined in the specification, where the
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC	of an oligonucleotide comprising a sequence complementary to the
CC	complementary strand of a polynucleotide which comprises a 5'-end
CC	sequence and an oligonucleotide comprising a sequence complementary to a
CC	polynucleotide which comprises a 3'-end sequence, where the
CC	oligonucleotide comprises at least 15 nucleotides and the combination of
CC	the 5'-end sequence/3'-end sequence is selected from those defined in
CC	the specification. The primer sets can be used in antisense therapy and
CC	in gene therapy. The primers are useful for synthesising polynucleotides,
CC	particularly full-length cDNAs. The primers are also useful for the
CC	detection and/or diagnosis of the abnormality of the proteins encoded by
CC	the full-length cDNAs. The primers allow obtaining of the full-length
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC	represent oligonucleotides, all of which are used in the exemplification
CC	of the present invention.
XX	
SQ	Sequence 531 AA:
Query Match	42.2%; Score 49; DB 22; Length 531;
Best Local Similarity	52.9%; Pred. No. 17;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0	
GY	5 LSTLEKMSNDPOYKND 21
:	: :   :
Dd	247 Invlagasefdpqykd 263
RESULT 9	
AAW88707	
ID	AAW88707 standard; Protein: 623 AA.
XX	
AC	AAW88707;
XX	
DT	01-MAR-1999 (first entry)
XX	
DE	Secreted protein encoded by gene 174 clone HE9PB42.
XX	
KX	Human; secreted protein; fusion protein; gene therapy; protein therapy;

KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia
KW	developmental abnormality; foetal deficiency; blood; allergy; renal;
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW	inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX	
OS	Homo sapiens.
XX	
PN	MO9854963-A2.
XX	
PD	10-DEC-1998.
XX	
PF	04-JUN-1998; 98MO-US11422.
XX	
PR	18-DEC-1997; 97US-0070923.
PR	06-JUN-1997; 97US-0048877.
PR	06-JUN-1997; 97US-0048881.
PR	06-JUN-1997; 97US-0048884.
PR	06-JUN-1997; 97US-0048883.
PR	06-JUN-1997; 97US-0048886.
PR	06-JUN-1997; 97US-0048899.
PR	06-JUN-1997; 97US-0048915.
PR	06-JUN-1997; 97US-0048949.
PR	06-JUN-1997; 97US-0048964.
PR	06-JUN-1997; 97US-0048972.
PR	06-JUN-1997; 97US-0049020.
PR	06-JUN-1997; 97US-0049375.
PR	05-SEP-1997; 97US-0057628.
PR	05-SEP-1997; 97US-0057635.
PR	05-SEP-1997; 97US-0057644.
PR	05-SEP-1997; 97US-0057647.
PR	05-SEP-1997; 97US-0057650.
PR	05-SEP-1997; 97US-0057661.
PR	05-SEP-1997; 97US-0057667.
PR	05-SEP-1997; 97US-0057761.
PR	05-SEP-1997; 97US-0057764.
PR	05-SEP-1997; 97US-0057770.
PR	05-SEP-1997; 97US-0057775.
PR	05-SEP-1997; 97US-0057778.
PR	06-JUN-1997; 97US-0048875.
PR	06-JUN-1997; 97US-0048878.
PR	06-JUN-1997; 97US-0048882.
PR	06-JUN-1997; 97US-0048885.
PR	06-JUN-1997; 97US-0048894.
PR	06-JUN-1997; 97US-0048897.
PR	06-JUN-1997; 97US-0048900.
PR	06-JUN-1997; 97US-0048916.
PR	06-JUN-1997; 97US-0048962.
PR	06-JUN-1997; 97US-0048970.
PR	06-JUN-1997; 97US-0048974.
PR	06-JUN-1997; 97US-0049373.
PR	05-SEP-1997; 97US-0057584.
PR	05-SEP-1997; 97US-0057629.
PR	05-SEP-1997; 97US-0057642.
PR	05-SEP-1997; 97US-0057645.
PR	05-SEP-1997; 97US-0057648.
PR	05-SEP-1997; 97US-0057651.
PR	05-SEP-1997; 97US-0057662.
PR	05-SEP-1997; 97US-0057668.
PR	05-SEP-1997; 97US-0057762.
PR	05-SEP-1997; 97US-0057765.
PR	05-SEP-1997; 97US-0057771.
PR	05-SEP-1997; 97US-0057776.
PR	06-JUN-1997; 97US-0048876.
PR	06-JUN-1997; 97US-0048880.
PR	06-JUN-1997; 97US-0048883.
PR	06-JUN-1997; 97US-0048882.
PR	06-JUN-1997; 97US-0048895.
PR	06-JUN-1997; 97US-0048898.
PR	06-JUN-1997; 97US-0048901.
PR	06-JUN-1997; 97US-0048917.



RESULT 11  
AAG09941  
ID AAG09941 standard: Protein; 195 AA.  
XX  
AC AAG09941;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8067.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
XX 06-SEP-2000.  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132653.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.



PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160983.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 40.5%; Score 47; DB 21; Length 195;  
Best Local Similarity 46.2%; Pred. NO. 10;  
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 2 INHSLLEAWSN 14  
Db 108 lnhgvidawsch 120

RESULT 12  
AAG31391  
ID AAG31391 standard; Protein: 371 AA.  
XX  
AC AAG31391;  
XX  
DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 37689.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
PN  
XX  
PD  
XX  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.

```

PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141847.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

```

```

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 39.7%; Score 46; DB 21; Length 371;
Best Local Similarity 52.9%; Pred. No. 33;
Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

Oy 7 IEAWSDN--DPOYKMD 21
  ||:|:| || ||
Db 166 lsswtengygdpsenkmd 182

RESULT 13
AAG31390
ID AAG31390 standard; Protein; 372 AA.
XX
AC AAG31390;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37688.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX

```

OS Arabidopsis thaliana.  
XX EPI033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0127462.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147038.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149425.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.



PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147203.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 39.7%; Score 46; DB 21; Length 411;  
Best Local Similarity 52.9%; Pred. No. 37;  
Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 7 ILEANSNDN-DPOYNKD 21  
||:::| || |||  
Db 166 llswtlenygdpsehk 182

RESULT 15  
AAG13577  
ID AAG13577 standard; Protein; 412 AA.

XX AAG13577;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SPQ ID NO: 13123.

XX Protein identification: signal transduction pathway; metabolic pathway;  
KW hydridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

OS EPI033405-A2.

XX EPI033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PF 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130047.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 01-SEP-1999; 99US-0151939.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153738.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159328.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Query Match

Best Local Similarity 39.7%; Score 46; DB 21; Length 412;  
Matches 9; Conservativity 52.9%; Pred. No. 38;

Mismatches 3; Indels 2; Gaps 1;

QY 7 ILEAMSDN--DPQYKND 21  
||:|:| || |||  
Db 167 IISWLENYGDPSNKD 183

Search completed: March 27, 2002, 13:57:55  
Job time: 523 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:29 ; Search time 87.3 Seconds

(without alignments)  
5.413 Million cell updates/sec

Title: US-09-290-049a-11

Perfect score: 116  
Sequence: 1 AINHLSTLEAWSNDNDPOYNKD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

1 number of hits satisfying chosen parameters: 212252

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PCRNUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	1430	3 US-09-008-172-2	Sequence 2, Appl
2	116	100.0	1430	4 US-09-210-361-6	Sequence 6, Appl
3	90	77.6	1577	2 US-08-793-824-2	Sequence 2, Appl
4	75	64.7	1475	3 US-09-007-999-2	Sequence 2, Appl
5	75	64.7	1475	4 US-09-210-361-2	Sequence 2, Appl
6	66	56.9	1375	4 US-09-210-361-4	Sequence 4, Appl
7	46	39.7	703	4 US-09-422-669-27	Sequence 27, Appl
8	45	38.8	324	3 US-08-508-761B-31	Sequence 31, Appl
9	45	38.8	657	3 US-08-508-761B-2	Sequence 2, Appl
10	44.5	38.4	259	1 US-08-277-311A-3	Sequence 3, Appl
11	44.5	38.4	259	2 US-08-473-750-6	Sequence 6, Appl
12	44.5	38.4	259	3 US-08-477-326-6	Sequence 6, Appl
13	43.5	37.5	197	2 US-08-897-020-7	Sequence 6, Appl
14	42	36.2	205	2 US-08-684-024-6	Sequence 6, Appl
15	42	36.2	205	3 US-09-145-668-6	Sequence 6, Appl
16	42	36.2	1001	1 US-07-797-556-6	Sequence 6, Appl
17	42	36.2	1001	1 US-07-943-843-2	Sequence 2, Appl
18	42	36.2	1001	1 US-08-347-003-2	Sequence 2, Appl
19	42	36.2	1097	1 US-07-943-843-6	Sequence 6, Appl
20	42	36.2	1097	1 US-08-347-003-6	Sequence 6, Appl
21	41.5	35.8	720	4 US-09-296-284-25	Sequence 25, Appl
22	41.5	35.8	754	4 US-09-296-284-4	Sequence 4, Appl
23	41	35.3	703	2 US-08-835-099A-1	Sequence 1, Appl
24	41	35.3	703	3 US-08-157-349-1	Sequence 1, Appl
25	41	35.3	712	2 US-08-835-099A-2	Sequence 1, Appl
26	41	35.3	712	3 US-09-157-349-2	Sequence 2, Appl
27	40.5	34.9	878	2 US-08-708-541A-26	Sequence 26, Appl

28	40.5	34.9	2186	2 US-08-822-445-2	Sequence 2, Appl
29	40.5	34.9	3672	2 US-08-822-445-12	Sequence 12, Appl
30	40.5	34.9	3801	2 US-08-822-445-10	Sequence 10, Appl
31	40	34.5	13	2 US-08-592-406-20	Sequence 20, Appl
32	40	34.5	119	1 US-08-340-539A-18	Sequence 18, Appl
33	40	34.5	280	1 US-08-434-255-8	Sequence 8, Appl
34	40	34.5	280	1 US-08-459-967-8	Sequence 8, Appl
35	40	34.5	280	1 US-08-460-327-8	Sequence 8, Appl
36	40	34.5	280	1 US-08-459-871-8	Sequence 8, Appl
37	40	34.5	280	4 US-09-024-532-2	Sequence 2, Appl
38	40	34.5	280	4 US-09-104-623A-2	Sequence 2, Appl
39	40	34.5	306	4 US-09-457-046B-4	Sequence 4, Appl
40	40	34.5	340	5 PCT-US96-02331-13	Sequence 13, Appl
41	40	34.5	370	1 US-08-434-255-6	Sequence 6, Appl
42	40	34.5	370	1 US-08-459-967-6	Sequence 6, Appl
43	40	34.5	370	1 US-08-460-327-6	Sequence 6, Appl
44	40	34.5	370	1 US-08-459-871-6	Sequence 6, Appl
45	40	34.5	397	1 US-08-434-255-2	Sequence 2, Appl

## ALIGNMENTS

```

RESULT 1
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: NICHOLS, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match      100.0%; Score 116; DB 3; Length 1430;
Best Local Similarity 100.0%; Pred. No. 8.2e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AINHLSTLEAWSNDNDPOYNKD 21
   |||||
Db 495 AINHLSTLEAWSNDNDPOYNKD 515

RESULT 2
US-09-210-361-6
; Sequence 6, App: cation US/09210361
; Patent No. 6281479
; GENERAL INFORMATION:
; APPLICANT: NICHOLS, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07

```

EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-210-361-6

Query Match 100.0%; Score 116; DB 4; Length 1430;  
Best Local Similarity 100.0%; Pred. No. 8,2e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AINHLSTLEAWSNDNDPOYKND 21  
|||||  
495 AINHLSTLEAWSNDNDPOYKND 515

RESULT 3  
US-08-793-824-2  
Sequence 2, Application US/08793824  
Patent No. 5981838  
GENERAL INFORMATION:  
APPLICANT: Simpson, Christine Lynn  
APPLICANT: Jacques, Phillip Morrison  
APPLICANT: Jacques, Nicholas Anthony  
TITLE OF INVENTION: Genetic Manipulation of Plants to  
TITLE OF INVENTION: Increase Stored Carbohydrates  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Griffith Hack & Co  
STREET: Level 8, 168 Walker Street  
CITY: No. 5981838ch Sydney  
STATE: New South Wales  
COUNTRY: Australia  
ZIP: 2060  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/793,824  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PM7643  
FILING DATE: 24-AUG-1994  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 61 2 9957 5944  
TELEFAX: 61 2 957 6288  
TELEX: 26547  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1577 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus salivarius  
US-08-793-824-2

Query Match 77.6%; Score 90; DB 2; Length 1577;  
Best Local Similarity 85.7%; Pred. No. 1.9e-06;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 AINHLSTLEAWSNDNDPOYKND 21

|||||  
DB 588 AIAHSLTLEAWSNDNDPOYKND 608

RESULT 4  
US-09-007-999-2  
Sequence 2, Application US/09007999  
Patent No. 6087559  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starch and  
FILE REFERENCE: 0356D  
CURRENT APPLICATION NUMBER: US/09/007,999  
CURRENT FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-007-999-2

Query Match 64.7%; Score 75; DB 3; Length 1475;  
Best Local Similarity 71.4%; Pred. No. 0.00053;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

1 AINHLSTLEAWSNDNDPOYKND 21  
|:|||||  
DB 481 ANDHSLTLEAWSNDNDPTYLHD 501

RESULT 5  
US-09-210-361-2  
Sequence 2, Application US/09210361  
Patent No. 6284479  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starches and  
FILE REFERENCE: 0357CR  
CURRENT APPLICATION NUMBER: US/09/210,361  
CURRENT FILING DATE: 1998-12-11  
EARLIER APPLICATION NUMBER: 09/007,999  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/009,620  
EARLIER FILING DATE: 1998-01-20  
EARLIER APPLICATION NUMBER: 08/485,243  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 64.7%; Score 75; DB 4; Length 1475;  
Best Local Similarity 71.4%; Pred. No. 0.00053;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

1 AINHLSTLEAWSNDNDPOYKND 21  
|:|||||

Db 481 ANDHLSILEAMSDNDPYLHD 501

## RESULT 6

US-09-210-361-4  
; Sequence 4, Application US/09210361  
; Patent No. 6284479

## GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starches and  
FILE REFERENCE: 0357CR  
CURRENT APPLICATION NUMBER: US/09/210.361  
EARLIER FILING DATE: 1998-12-11  
EARLIER APPLICATION NUMBER: 09/007,999  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/009,620  
EARLIER FILING DATE: 1998-01-20  
EARLIER APPLICATION NUMBER: 08/485,243  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 1375  
TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-210-361-4

Query Match 56.9%; Score 66; DB 4; Length 1375;  
Best Local Similarity 66.7%; Pred. No. 0.015;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Oy 1 AINHLSTLEAMSDNDPOYNKD 21

Db 507 ANDHLSILEAMSYNDPYLHD 527

## RESULT 7

US-09-422-869-27  
; Sequence 27, Application US/09422869  
; Patent No. 6235481

## GENERAL INFORMATION:

APPLICANT: POLONSKY, KENNETH S.  
APPLICANT: HORIKAWA, YUKIO  
APPLICANT: ODA, NAOHISA  
APPLICANT: COX, NANCY J.  
APPLICANT: SREENAN, SEAMUS  
APPLICANT: ZHOU, YUN-PING  
APPLICANT: OTANI, KENICHI  
APPLICANT: HANIS, CRAIG L.  
APPLICANT: BELL, GRAEME I.  
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES  
FILE REFERENCE: ARCD:307  
CURRENT APPLICATION NUMBER: US/09/422,869  
CURRENT FILING DATE: 1999-10-21  
EARLIER APPLICATION NUMBER: 60/134,175  
EARLIER FILING DATE: 1999-05-13  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 27  
LENGTH: 703  
TYPE: PRT  
ORGANISM: RAT  
US-09-422-869-27

Query Match 39.7%; Score 46; DB 4; Length 703;  
Best Local Similarity 70.0%; Pred. No. 14;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 10 AMSDNDPOYN 19

Db 296 AMSDNAPEWN 305

## RESULT 8

US-08-508-761B-31  
; Sequence 31, Application US/08508761B  
; Patent No. 6027520

## GENERAL INFORMATION:

APPLICANT: Jolliff, Gwennael  
APPLICANT: Guyonvarch, Armel  
APPLICANT: Purification, Relano  
APPLICANT: Duchiron, Francis  
APPLICANT: Renaud, Michel  
TITLE OF INVENTION: System for Protein Expression and  
SECRETION ESPECIALLY IN CORYNEBACTERIA  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC  
STREET: 400 Seventh St. N.W.  
CITY: Washington D.C.  
COUNTRY: U.S.A.  
ZIP: 20004

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/508,761B

FILING DATE: 31-JUL-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 91/09652

FILING DATE: 29-JUL-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 91/09870

FILING DATE: 02-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Player, William E.

REGISTRATION NUMBER: 31,409

REFERENCE/DOCKET NUMBER: P58525NA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 638-6666

TELEFAX: (202) 393-5350

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:

LENGTH: 324 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: C. glutamicum

US-08-508-761B-31

Query Match 38.8%; Score 45; DB 3; Length 324;  
Best Local Similarity 53.8%; Pred. No. 8.2;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 9 EAMSDNDPOYNKD 21

Db 212 ERMQENDPKSNVD 224

## RESULT 9

US-08-508-761B-2  
; Sequence 2, Application US/08508761B

Patent No. 6027920  
GENERAL INFORMATION:  
APPLICANT: Joliff, Gwennael  
APPLICANT: Guyonvarch, Arnel  
APPLICANT: Purification, Relano  
APPLICANT: Duchilton, Francis  
TITLE OF INVENTION: System for Protein Expression and  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC  
STREET: 400 Seventh St. N.W.  
CITY: Washington D.C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/508,761B  
FILING DATE: 31-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91/09652  
FILING DATE: 29-JUL-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91/09870  
FILING DATE: 02-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Player, William E.  
REGISTRATION NUMBER: 31,409  
REFERENCE/DOCKET NUMBER: P58525NA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-6666  
TELEFAX: (202) 393-5350  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 657 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-508-761B-2

Query Match 38.8%; Score 45; DB 3; Length 657;  
Best Local Similarity 53.8%; Pred. No. 19;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 9 EAMSDNDPOYNKD 21  
| | :||| :| |  
DB 286 ERMOENDPKSNVD 298

RESULT 10  
US-08-277-231A-3  
Sequence 3, Application US/08277231A  
Patent No. 5643725  
GENERAL INFORMATION:  
APPLICANT: Green, Bruce A.  
APPLICANT: Brinton, Charles C.  
TITLE OF INVENTION: Sequence and Analysis of LKP Pilln  
Patent No. 5643725  
TITLE OF INVENTION: Structural Genes and The LKP Pill Operon of No. 5643725typable  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: U.S.A.

ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/277,231A  
FILING DATE: 19-JUL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Alice O.  
REGISTRATION NUMBER: 32,542  
REFERENCE/DOCKET NUMBER: ACC94-02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 259 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-277-231A-3

Query Match 38.4%; Score 44.5; DB 1; Length 259;  
Best Local Similarity 40.0%; Pred. No. 7.6;  
Matches 8; Conservative 5; Mismatches 2; Indels 5; Gaps 1;

QY 6 SILEAMSDN-----DPOYNK 20  
:::| | | | | | | | | |  
DB 76 ALYQAWIDNGNPNADPKYTK 95

RESULT 11  
US-08-473-750-6  
Sequence 6, Application US/08473750  
Patent No. 5834187  
GENERAL INFORMATION:  
APPLICANT: Green, Bruce A.  
APPLICANT: Brinton, Jr., Charles C.  
TITLE OF INVENTION: Sequence and Analysis of LKP Pilln  
Patent No. 5834187  
TITLE OF INVENTION: Structural Gene and the LKP Pill Operon of No. 5834187 5786  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,750  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/277,321  
FILING DATE: 19-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Alice O.  
REGISTRATION NUMBER: 33,542  
REFERENCE/DOCKET NUMBER: ACC94-02B  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 259 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-473-750-6

Query Match 38.4%; Score 44.5; DB 2; Length 259;  
Best Local Similarity 40.0%; Pred. No. 7.6;  
Matches 8; Conservative 5; Mismatches 2; Indels 5; Gaps 1;

OY 6 SILEAMSDN-----DPQNK 20  
:::|||||  
Db 76 ALVQAMIDNGNPNADPKYTK 95

ULT 12  
Sequence 6, Application US/08477326  
Patent No. 5968769  
GENERAL INFORMATION:  
APPLICANT: Brinton, Jr., Charles C.  
TITLE OF INVENTION: Sequence and Analysis of LKP P11n  
Patent No. 5968769  
TITLE OF INVENTION: Structural Gene and the LKP P11 Operon of No. 5968769typable  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,326  
FILING DATE: 07-JUN-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/277,231  
FILING DATE: July 19, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Alice O.  
REGISTRATION NUMBER: 33,542  
REFERENCE/DOCKET NUMBER: ACC94-02A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 259 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-477-326-6

Query Match 38.4%; Score 44.5; DB 2; Length 259;  
Best Local Similarity 40.0%; Pred. No. 7.6;  
Matches 8; Conservative 5; Mismatches 2; Indels 5; Gaps 1;  
OY 6 SILEAMSDN-----DPQNK 20  
:::|||||

Db 76 ALVQAMIDNGNPNADPKYTK 95

RESULT 13  
US-08-897-020-7  
Sequence 7, Application US/08897020  
Patent No. 6028176  
GENERAL INFORMATION:  
APPLICANT: Shanefelt, Armen; Greve, Jeffrey; Rocznia, Steven  
TITLE OF INVENTION: High-affinity Interleukin-4 Mutelins  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bayer Corporation, Pharmaceutical Division  
STREET: 400 Morgan Lane  
CITY: West Haven  
STATE: CT  
COUNTRY: United States of America  
ZIP: 06516-4175  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS v. 6.30  
SOFTWARE: Word for Windows 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/897,020  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P-91,242  
FILING DATE: 19-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Huw R. Jones  
REGISTRATION NUMBER: 33,916  
REFERENCE/DOCKET NUMBER: WH5020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (203) 812-2317  
TELEFAX: (203) 812-5492  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 197  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: SIL-4R-STX  
HYPOTHETICAL: no  
ANTI-SENSE: no  
US-08-897-020-7

Query Match 37.5%; Score 43.5; DB 3; Length 197;  
Best Local Similarity 53.3%; Pred. No. 8;  
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

OY 3 NHLS-ILEAMSDNP 16  
|||:|||||  
Db 110 NHLYAVNINWSENP 124

RESULT 14  
US-08-684-024-6  
Sequence 6, Application US/08684024  
Patent No. 5834298  
GENERAL INFORMATION:  
APPLICANT: Benzeira, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.

ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/684,024  
FILING DATE: 19-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-A  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-684-024-6

Query Match 36.2%; Score 42; DB 2; Length 205;  
Best Local Similarity 40.0%; Pred. No. 15;  
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 4 HLSLEAWSNDNDPOY 18  
||:|::||:  
DB 160 HLVEPEKWEESGPQF 174

RESULT 15  
US-09-145-868-6  
Sequence 6, Application US/09145868  
Patent No. 6096522  
GENERAL INFORMATION:  
APPLICANT: Benezira, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/145,868  
FILING DATE: 02-SEP-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-B  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-145-868-6

Query Match 36.2%; Score 42; DB 3; Length 205;  
Best Local Similarity 40.0%; Pred. No. 15;  
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 4 HLSLEAWSNDNDPOY 18  
||:|::||:  
DB 160 HLVEPEKWEESGPQF 174

Search completed: March 27, 2002, 13:59:30  
Job time: 583 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:24 ; Search time 1139.61 Seconds

(without alignments)  
5.116 Million cell updates/sec

Title: US-09-290-049a-11

Perfect score: 116

Sequence: 1 AINHLSTLEAWSNDPQYKND 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Maximum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US087.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US087.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US087.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US081.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US081.COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US083.COMB.pep.\*  
8: /cgn2\_6/ptodata/2/paa/US084.COMB.pep.\*  
9: /cgn2\_6/ptodata/2/paa/US085.COMB.pep.\*  
10: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
11: /cgn2\_6/ptodata/2/paa/US087.COMB.pep.\*  
12: /cgn2\_6/ptodata/2/paa/US088.COMB.pep.\*  
13: /cgn2\_6/ptodata/2/paa/US089.COMB.pep.\*  
14: /cgn2\_6/ptodata/2/paa/US090.COMB.pep.\*  
15: /cgn2\_6/ptodata/2/paa/US091.COMB.pep.\*  
16: /cgn2\_6/ptodata/2/paa/US092.COMB.pep.\*  
17: /cgn2\_6/ptodata/2/paa/US093.COMB.pep.\*  
18: /cgn2\_6/ptodata/2/paa/US094.COMB.pep.\*  
19: /cgn2\_6/ptodata/2/paa/US095.COMB.pep.\*  
20: /cgn2\_6/ptodata/2/paa/US096.COMB.pep.\*  
21: /cgn2\_6/ptodata/2/paa/US097.COMB.pep.\*  
22: /cgn2\_6/ptodata/2/paa/US098.COMB.pep.\*  
23: /cgn2\_6/ptodata/2/paa/US099.COMB.pep.\*  
24: /cgn2\_6/ptodata/2/paa/US099.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	100.0	21	US-09-290-049-11	Sequence 11, Appl
2	116	100.0	1430	US-09-649-885-2	Sequence 2, Appl
3	116	100.0	1430	US-09-740-274-6	Sequence 6, Appl
4	75	64.7	21	US-09-290-049-1	Sequence 1, Appl
5	75	64.7	21	US-09-290-049-10	Sequence 10, Appl
6	75	64.7	21	US-09-290-049-13	Sequence 13, Appl
7	75	64.7	1475	US-09-557-848-2	Sequence 2, Appl
8	75	64.7	1475	US-09-740-274-12	Sequence 2, Appl
9	74	63.8	21	US-09-290-049-12	Sequence 12, Appl

10	74	63.8	21	US-09-290-049-14	Sequence 14, Appl
11	66	56.9	1375	US-09-740-274-4	Sequence 4, Appl
12	64	55.2	2057	US-09-499-203-2	Sequence 2, Appl
13	49	42.2	623	PCT-US01-05614-482	Sequence 482, App
14	49	42.2	623	PCT-US98-11422A-407	Sequence 407, App
15	49	42.2	623	US-09-205-258-482	Sequence 482, App
16	49	42.2	623	US-09-933-767-482	Sequence 482, App
17	49	42.2	704	PCT-US01-05614-422	Sequence 422, App
18	49	42.2	704	US-09-205-258-422	Sequence 422, App
19	49	42.2	704	US-09-933-767-422	Sequence 422, App
20	48	41.4	883	US-09-489-039A-9716	Sequence 9716, App
21	47	40.5	196	US-09-595-298A-765	Sequence 765, App
22	47	40.5	521	US-09-107-532-6431	Sequence 6431, App
23	47	40.5	521	US-09-107-532A-6431	Sequence 6431, App
24	46.5	40.1	327	US-09-107-532-6181	Sequence 6181, App
25	46.5	40.1	327	US-09-107-532A-6181	Sequence 6181, App
26	46	39.7	199	US-09-391-631-2917	Sequence 2917, App
27	46	39.7	247	US-09-391-631-2916	Sequence 2916, App
28	46	39.7	647	US-09-281-253-10	Sequence 10, Appl
29	46	39.7	647	US-09-281-253-10	Sequence 10, Appl
30	46	39.7	647	US-09-281-253-10	Sequence 10, Appl
31	46	39.7	703	US-09-768-877-27	Sequence 10, Appl
32	46	39.7	1652	US-09-242-679-1100	Sequence 1100, Ap
33	46	39.7	10182	US-09-134-001C-3159	Sequence 3159, Ap
34	46	39.7	10203	US-09-450-969-4098	Sequence 4098, Ap
35	45	38.8	151	US-09-198-452A-815	Sequence 815, App
36	45	38.8	205	US-09-438-185-767	Sequence 767, App
37	45	38.8	365	US-09-312-344-8416	Sequence 8416, App
38	45	38.8	565	US-09-602-839A-534	Sequence 534, App
39	45	38.8	657	US-08-039-028A-2	Sequence 2, Appl
40	45	38.8	657	US-08-039-028A-2	Sequence 2, Appl
41	45	38.8	657	US-08-039-028A-2	Sequence 2, Appl
42	45	38.8	657	US-08-508-761-2	Sequence 2, Appl
43	45	38.8	657	US-09-738-626-6670	Sequence 6670, Ap
44	45	38.8	790	US-09-543-681A-6460	Sequence 6460, Ap
45	44.5	38.4	259	US-08-277-231-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-09-290-049-11  
Sequence 11, Application US/09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: FDC98-0192A  
CURRENT FILING DATE: 1999-04-12  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 11  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. mutans  
US-09-290-049-11

Query Match 100.0%; Score 116; DB 16; Length 21;  
Best Local Similarity 100.0%; Pred. No. 4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AINHLSTLEAWSNDPQYKND 21  
Db 1 AINHLSTLEAWSNDPQYKND 21

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

Query Match Similarity      100.0%; Score 116; DB 20; Length 1430;
Best Local Similarity      100.0%; Pred. No. 4.4e-09;
Matches      21; Conservative      0; Mismatches      0; Indels      0; Gaps      0.

OY      1 A1NHLSTLEAWSNDPNQYKND 21
      |||||
Db      495 A1NHLSTLEAWSNDPNQYKND 515

RESULT      3
US-09-740-274-6
; Sequence 6, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: NICHOLS, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/220,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

```

```

RESULT      4
US-09-290-049-1
: Sequence 1, Application US/09290049
: GENERAL INFORMATION:
:   APPLICANT: Smith, Daniel J.
:   APPLICANT: Taubman, Martin A.
:   TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
:   TITLE OF INVENTION: CARIES
:   FILE REFERENCE: PDC98-01P2A
:   CURRENT APPLICATION NUMBER: US/09/290, 049
:   CURRENT FILING DATE: 1999-04-12
:   EARLIER APPLICATION NUMBER: 60/081,550
:   EARLIER FILING DATE: 1998-04-13
:   EARLIER APPLICATION NUMBER: 60/113,142
:   EARLIER FILING DATE: 1999-01-08
:   NUMBER OF SEQ ID NOS: 19
:   SOFTWARE: FastSeq for Windows Version 3.0
:   SEQ ID NO 1
:   LENGTH: 21
:   TYPE: PRT
:   ORGANISM: Artificial Sequence
:   FEATURE:
:   OTHER INFORMATION: EAW peptide
US-09-290-049-1

```

	Query Match	Similarity	64.7%	Score	75	DB	16	Length	21												
	Best Local	Similarity	71.4%	Pred.	No.	0.00016															
	Matches	15;	Conservative	1;	Mismatches	5;	Indels	0;	Gaps	0;											
Oy	1	A	I	N	H	L	S	T	E	A	M	S	D	N	D	P	O	K	N	D	21
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	1	A	N	D	H	L	S	T	E	A	M	S	D	N	D	P	O	K	N	D	21

```

RESULT      5
US-09-290-049-10
: Sequence 10, Application US/09290049
: GENERAL INFORMATION:
: APPLICANT: Smith, Daniel J.
: APPLICANT: Taubman, Martin A.
: TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
: TITLE OF INVENTION: CARRIES
: FILE REFERENCE: EDC98-01P2A
: CURRENT APPLICATION NUMBER: US/09/290,049
: CURRENT FILING DATE: 1999-04-12
: EARLIER APPLICATION NUMBER: 60/081,550
: EARLIER FILING DATE: 1998-04-13
: EARLIER APPLICATION NUMBER: 60/115,142
: EARLIER FILING DATE: 1999-01-08
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 10
: LENGTH: 21
: TYPE: PRT
: ORGANISM: S. mutans
US-09-290-049-10

```

Query Match	64.7%	Score 75	DB 16	Length 21
Best Local Similarity	71.4%	Pred. No. 0.00016		
Matches 15: Conservative	1:	Mismatches 5:	Indels 0:	Gaps 0:

	Query Match	100.0%	Score 116;	DB 21;	Length 1430;
	Best Local Similarity	100.0%;	Pred. No.	4.4e-09;	
	Matches    21; Conservative	0;	Mismatches	0;	Indels      0;
Oy	1 AINHLSTLEAWSNDNDPOYNKD	21			
Db	495 AINHLSITLAWSDNDPOYNKD	515			

```

Oy      1 AINHLSTLEAWSDNDPOYKND 21
        | : ||||| ||||| || |
Db      1 ANDHLSITLEAWSDNDPTYLND 21

RESULT      6
US-09-290-049-13
; Sequence 13, Application US/09290049
; GENERAL INFORMATION:

```

APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
CRIES  
FILE REFERENCE: FDC98-01D2A  
CURRENT APPLICATION NUMBER: US/09/290,049  
CURRENT FILING DATE: 1999-04-12  
EARLIER APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115,142  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 13  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. downei  
US-09-290-049-13

Query Match 64.7%; Score 75; DB 16; Length 21;  
Best Local Similarity 75.0%; Pred. No. 0.00016;  
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 AINHLSTLEAWSNDNDPOYNK 20  
DB 1 AIDHLSILEAWSNDNDPYLHD 20

RESULT 7  
US-09-557-848-2  
Sequence 2, Application US/09557848  
GENERAL INFORMATION:  
APPLICANT: Nicholas, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starch and  
FILE REFERENCE: 0356D2  
CURRENT APPLICATION NUMBER: US/09/557,848  
CURRENT FILING DATE: 2000-04-26  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/007,999  
EARLIER FILING DATE: 1998-01-16  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-557-848-2

Query Match 64.7%; Score 75; DB 19; Length 1475;  
Best Local Similarity 71.4%; Pred. No. 0.018;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHLSTLEAWSNDNDPOYNK 21  
DB 481 ANDHLSILEAWSNDNDPYLHD 501

RESULT 8  
US-09-740-274-2  
Sequence 2, Application US/09740274  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Glucan-containing Compositions and Paper  
FILE REFERENCE: 0357CRD  
CURRENT APPLICATION NUMBER: US/09/740,274  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 09/210,361  
PRIOR FILING DATE: 1998-12-11  
PRIOR APPLICATION NUMBER: 09/007,999

PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 64.7%; Score 75; DB 21; Length 1475;  
Best Local Similarity 71.4%; Pred. No. 0.018;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHLSTLEAWSNDNDPOYNK 21  
DB 481 ANDHLSILEAWSNDNDPYLHD 501

RESULT 9  
US-09-290-049-12  
Sequence 12, Application US/09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
CRIES  
FILE REFERENCE: FDC98-01D2A  
CURRENT APPLICATION NUMBER: US/09/290,049  
CURRENT FILING DATE: 1999-04-12  
EARLIER APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115,142  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. downei  
US-09-290-049-12

Query Match 63.8%; Score 74; DB 16; Length 21;  
Best Local Similarity 66.7%; Pred. No. 0.00023;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHLSTLEAWSNDNDPOYNK 21  
DB 1 ANNHVSIWEAWSNDNDPYLHD 21

RESULT 10  
US-09-290-049-14  
Sequence 14, Application US/09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
CRIES  
FILE REFERENCE: FDC98-01D2A  
CURRENT APPLICATION NUMBER: US/09/290,049  
CURRENT FILING DATE: 1999-04-12

EARLIER APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115,142  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 14  
LENGTH: 21  
TYPE: PRF  
ORGANISM: S. sobrinus  
US-09-290-049-14

Query Match 63.8%; Score 74; DB 16; Length 21;  
Best Local Similarity 66.7%; Pred. No. 0.00023;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSTLEWMSDNDPQYKND 21  
1 ANNHVSIVEAWMSDNDPPLDHD 21

RESULT 11  
US-09-740-274-4  
Sequence 4, Application US/09740274  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Glucan-containing Compositions and Paper  
FILE REFERENCE: 0357CARD  
CURRENT APPLICATION NUMBER: US/09/740,274  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 09/210,361  
PRIOR FILING DATE: 1998-12-11  
PRIOR APPLICATION NUMBER: 09/007,999  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 1375  
TYPE: PRF  
ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 56.9%; Score 66; DB 21; Length 1375;  
Best Local Similarity 66.7%; Pred. No. 0.46;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 AINHLSTLEWMSDNDPQYKND 21  
Db 507 ANDHLSTLEWMSDNDPPLDHD 527

RESULT 12  
US-09-499-203-2  
Sequence 2, Application US/09499203  
GENERAL INFORMATION:  
APPLICANT: KOSSMANN, Jens  
APPLICANT: WELSH, Thomas  
APPLICANT: QUANZ, Martin  
APPLICANT: KNUTH, Karola  
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
FILE REFERENCE: 147-196P

CURRENT APPLICATION NUMBER: US/09/499,203  
CURRENT FILING DATE: 2000-02-08  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 2057  
TYPE: PRF  
ORGANISM: Leuconostoc mesenteroides  
US-09-499-203-2

Query Match 55.2%; Score 64; DB 18; Length 2057;  
Best Local Similarity 66.7%; Pred. No. 1.5;  
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSTLEWMSDNDPQY 18  
Db 665 ANKHLSTLEWMSDNDPQY 682

RESULT 13  
PCT-US01-05614-482  
Sequence 482, Application PC/TUS0105614  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc.  
TITLE OF INVENTION: 207 Human Secreted Proteins  
FILE REFERENCE: P2007PCT2  
CURRENT APPLICATION NUMBER: PCT/US01/05614  
CURRENT FILING DATE: 2001-02-21  
PRIOR APPLICATION NUMBER: 60/184,836  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/193,170  
PRIOR FILING DATE: 2000-03-29  
NUMBER OF SEQ ID NOS: 1245  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 482  
LENGTH: 623  
TYPE: PRF  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (111)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (575)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
PCT-US01-05614-482

Query Match 42.2%; Score 49; DB 1; Length 623;  
Best Local Similarity 52.9%; Pred. No. 1e+02;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 LSTLEWMSDNDPQYKND 21  
Db 107 IWLXGASEFDPOYKND 123

RESULT 14  
PCT-US98-11422A-407  
Sequence 407, Application PC/TUS9811422A  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc., et al.  
TITLE OF INVENTION: 207 Human Secreted Proteins  
NUMBER OF SEQUENCES: 800  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.,  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US98/11422A  
FILING DATE: 04-JUN-1998  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 407:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 623 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US98-11422A-407

Query Match 42.2%; Score 49; DB 1; Length 623;  
Best Local Similarity 52.9%; Pred. No. 1e+02;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 LSLLEAWSNDPOYKND 21  
: : : : :  
Db 107 IMVLGASEFPDQYKND 123

RESULT 15  
US-09-205-258-482  
Sequence 482, Application US/09205258  
GENERAL INFORMATION:  
APPLICANT: Young et al.  
FILE REFERENCE: P2007P1  
CURRENT APPLICATION NUMBER: US/09/205,258  
CURRENT FILING DATE: 1998-12-04  
EARLIER APPLICATION NUMBER: PCT/US98/11422  
EARLIER FILING DATE: 1998-06-04  
EARLIER APPLICATION NUMBER: 60/048,885  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,375  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,881  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,880  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,896  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,020  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,876  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,895  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,884  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,894  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,971  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,882  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,899

EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,893  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,900  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,901  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,892  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,915  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,019  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,970  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,972  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,916  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,373  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,875  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,374  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,917  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,949  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,883  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,897  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,898  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,962  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,963  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,877  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,878  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/070,923  
EARLIER FILING DATE: 1997-12-18  
EARLIER APPLICATION NUMBER: 60/092,921  
EARLIER FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/094,657  
EARLIER FILING DATE: 1998-07-30  
NUMBER OF SEQ ID NOS: 1227  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 482  
LENGTH: 623  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (111)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (575)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
US-09-205-258-482

Query Match 42.2%; Score 49; DB 16; Length 623;  
Best Local Similarity 52.9%; Pred. No. 1e+02;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 LSLLEAWSNDPOYKND 21

Db 107 IMVLXGASEFDPQYNKD 123

Search completed: March 27, 2002, 14:20:25  
Job time: 1572 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:47 ; Search time 137.48 Seconds  
(without alignments)  
10,540 Million cell updates/sec

Title: US-09-290-049A-11

Perfect score: 116

Sequence: 1 AINHLSTLEAWSNDNDPOYKND 21

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

332938 seqs, 6899538 residues

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCF\_NEW.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW.COMB.pep1:\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW.COMB.pep1:\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	100.0	21	US-09-562-328-22	Sequence 22, Appl
2	116	100.0	21	US-09-290-049A-11	Sequence 11, Appl
3	116	100.0	545	US-09-604-957-4	Sequence 4, Appl
4	75	64.7	21	US-09-562-328-20	Sequence 20, Appl
5	75	64.7	21	US-09-562-328-24	Sequence 24, Appl
6	75	64.7	21	US-09-290-049A-1	Sequence 1, Appl
7	75	64.7	21	US-09-290-049A-10	Sequence 10, Appl
8	75	64.7	21	US-09-290-049A-13	Sequence 13, Appl
9	74	63.8	21	US-09-562-328-23	Sequence 23, Appl
10	74	63.8	21	US-09-562-328-25	Sequence 25, Appl
11	74	63.8	21	US-09-290-049A-12	Sequence 12, Appl
12	74	63.8	21	US-09-290-049A-14	Sequence 14, Appl
13	68	58.6	523	US-09-604-957-5	Sequence 5, Appl
14	66	56.9	21	US-09-562-328-21	Sequence 21, Appl
15	64	55.2	584	US-09-604-957-6	Sequence 6, Appl
16	53.5	46.1	535	US-09-604-957-7	Sequence 7, Appl
17	53.5	46.1	1278	US-09-604-957-3	Sequence 3, Appl
18	51	44.0	429	PCT-US02-03987-14022	Sequence 14022, A
19	51	44.0	429	US-09-815-242-14022	Sequence 14022, A
20	51	44.0	429	US-09-815-242-14022	Sequence 14022, A
21	49	42.2	623	US-10-023-282-482	Sequence 482, App
22	49	42.2	704	US-10-023-282-422	Sequence 422, App
23	49	42.2	704	US-09-611-526-3841	Sequence 3841, App
24	47	40.5	95	US-09-945-301-14	Sequence 14, Appl
25	46	39.7	371	US-09-708-427-29185	Sequence 29185, A

26	46	39.7	372	6	US-09-708-427-29184	Sequence 29184, A
27	46	39.7	411	6	US-09-620-394B-1367	Sequence 1367, Ap
28	46	39.7	412	6	US-09-620-394B-1366	Sequence 1366, Ap
29	46	39.7	420	6	US-09-708-427-29183	Sequence 29183, A
30	46	39.7	460	6	US-09-620-394B-1365	Sequence 1365, Ap
31	45	38.8	330	6	US-09-708-427-72864	Sequence 72864, A
32	45	38.8	364	6	US-09-708-427-72863	Sequence 72863, A
33	45	38.8	376	6	US-09-708-427-72862	Sequence 72862, A
34	45	38.8	596	6	US-09-708-427-31640	Sequence 31640, A
35	45	38.8	597	6	US-09-708-427-31639	Sequence 31639, A
36	45	38.8	637	6	US-09-708-427-31638	Sequence 31638, A
37	44	37.9	869	6	US-09-614-150-423	Sequence 423, App
38	44	37.9	1017	6	US-09-614-150-35226	Sequence 35226, A
39	43.5	37.5	691	6	US-09-313-942-18	Sequence 20, Appl
40	43.5	37.5	694	6	US-09-313-942-18	Sequence 18, Appl
41	43.5	37.5	694	6	US-09-313-942-22	Sequence 22, Appl
42	43.5	37.5	784	6	US-09-313-942-30	Sequence 30, Appl
43	43.5	37.5	789	6	US-09-611-526-2651	Sequence 2651, Ap
44	43.5	37.5	793	6	US-09-313-942-32	Sequence 32, Appl
45	43.5	37.5	825	7	US-10-010-802-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-09-562-328-22  
; Sequence 22, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 0495, 0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-22

Query Match 100.0%; Score 116; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.3e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AINHLSTLEAWSNDNDPOYKND 21  
DB 1 AINHLSTLEAWSNDNDPOYKND 21  
RESULT 2  
US-09-290-049A-11  
; Sequence 11, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564, 1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0

```

; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049A-11

```

Query Match	100.0%;	Score 116;	DB 6;	length 21;
Best Local Similarity	100.0%;	Pred. No. 9.3e-12;		
Matches 21; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

```
QY 1 AINHLSILEAWSNDPQYNKD 21
    | | | | | | | | | | | | |
Db 1 AINHLSILEAWSNDPQYNKD 21
```

RESULT 3  
US-09-604-957-4  
Sequence 4, Application US/09604957  
GENERAL INFORMATION:

APPLICANT: VAN GELT-SCHUTTEN, GERTRIDINA HENDRIKA  
APPLICANT: DIFKHUIZEN, LUBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 4338  
CURRENT APPLICATION NUMBER: US/09/604,957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 545  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-604-957-4

Query Match	100.0%;	Score 116;	DB 6;	Length 545;
Best Local Similarity	100.0%;	Pred. No. 3.9e-10;		
Matches 21; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0

```
QY      1 AINHLSILEAWSNDNPQYNKD 21
        |||||
Db      75 AINHLSILEAWSNDNPQYNKD 95
```

[illegible]

Query Match	64.78;	Score 75;	DB 6;	Length 21;
Best Local Similarity	71.48;	Pred. No. 2.4e-05;		
Matches 15; Conservative	1;	Mismatches 5;	Indels 0;	Gaps 0

```
QY 1 AINHLSILEAWSNDNDPQYNKD 21
      | : ||||| ||||| | |
Db 1 ANDHLSILEAWSNDNDTPYLHD 21
```

RESULT 5  
US-09-562-328-24  
Sequence 24, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995, 0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 24  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-24

Query Match	64.7%	Score 75	DB 6	Length 21
Best Local Similarity	75.0%	Pred. No. 2.4e-05		
Matches	15	Conservative	1	Mismatches 4
				Indels 0
				Gaps 0

```
QY 1 AINHLSTLEAWSNDNDPQYNK 20
    ||:|||||  ||
Db 1 AIDHLSILEAWSGNDNDYVK 20
```

```

RESULT      6
US-09-290-049A-1
Sequence 1: Application US/09290049A
GENERAL INFORMATION:
APPLICANT: Smith, Daniel J.
APPLICANT: Taubman, Martin A.
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
TITLE OF INVENTION: CARIES
FILE REFERENCE: 1564,1008-002
CURRENT APPLICATION NUMBER: US/09/290,049A
CURRENT FILING DATE: 1999-04-12
PRIOR APPLICATION NUMBER: 60/081,550
PRIOR FILING DATE: 1998-04-13
PRIOR APPLICATION NUMBER: 60/115,142
PRIOR FILING DATE: 1999-01-08
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 21
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: EAW pep tide
US-09-290-049A-1

```

Query Match	64.78;	Score 75;	DB 6;	Length 21;
Best Local Similarity	71.48;	Pred. No. 2.4e-05;		
Matches 15;	Conservative 1;	Mismatches 5;	Indels 0;	Gaps 0

QY 1 AINHLILEAWSNDNPQYNKD 21  
| : ||||| |  
Db 1 ANDHLSILEAWSNDNTPYLHD 21

## RESULT 7



```

US-09-290-049A-10
; Sequence 10, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049A-13

RESULT 8
US-09-290-049A-13
; Sequence 13, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049A-13

Query Match      64.7%: Score 75; DB 6; Length 21;
Best Local Similarity 75.0%: Pred. No. 2.4e-05;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AINHLSTLEAWSNDNPQYNK 21
   | : ||||| ||||| | |
DB 1 ANDHLSILEAWSGNDNTPYLHD 21

RESULT 9
US-09-562-328-23
; Sequence 23, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES

```

```

FILE REFERENCE: 04995.0046-01
CURRENT APPLICATION NUMBER: US/09/562.328
CURRENT FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 09/288,965
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 23
LENGTH: 21
TYPE: PRT
ORGANISM: Streptococcus sp.
US-09-562-328-23

Query Match      63.8%, Score 74; DB 6; Length 21,
Best Local Similarity 66.7%; Pred.No. 3.4e-05;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY      1 AINHLSTLEAWSNDNDPOYNKD 21
        ||::||::||||| 1 |
Db       1 ANNHVSIVEAWSNDNDTPYLHD 21

RESULT 10
US-09-562-328-25
Sequence 25, Application US/09562328
GENERAL INFORMATION:
APPLICANT: LEEB, ANDREW
APPLICANT: TAUBMAN, MARTIN A.
APPLICANT: SMITH, DANIEL J.
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES
FILE REFERENCE: 04995.0046-01
CURRENT APPLICATION NUMBER: US/09/562.328
CURRENT FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 09/288,965
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 25
LENGTH: 21
TYPE: PRT
ORGANISM: Streptococcus sp.
US-09-562-328-25

Query Match      63.8%, Score 74; DB 6; Length 21;
Best Local Similarity 66.7%; Pred.No. 3.4e-05;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY      1 AINHLSTLEAWSNDNDPOYNKD 21
        ||::||::||||| 1 |
Db       1 ANNHVSIVEAWSNDNDTPYLHD 21

RESULT 11
US-09-290-049A-12
Sequence 12, Application US/09290049A
GENERAL INFORMATION:
APPLICANT: Smith, Daniel J.
APPLICANT: Taubman, Martin A.
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
FILE REFERENCE: 1564.1008-002
CURRENT APPLICATION NUMBER: US/09/290.049A
CURRENT FILING DATE: 1999-04-12
PRIOR APPLICATION NUMBER: 60/081,550
PRIOR FILING DATE: 1998-04-13
PRIOR APPLICATION NUMBER: 60/115,142
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 21

```

TYPE: PRT  
ORGANISM: S. downei  
US-09-290-049A-12

Query Match 63.8%; Score 74; DB 6; Length 21;  
Best Local Similarity 66.7%; Pred. No. 3.4e-05;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDNDPOYKND 21  
DB 1 ANNHVSIVEAWSNDNDPTPLHD 21

RESULT 12  
US-09-290-049A-14  
Sequence 14, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: 1564.1008-002  
CURRENT APPLICATION NUMBER: US/09/290.049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. sobrinus  
US-09-290-049A-14

Query Match 63.8%; Score 74; DB 6; Length 21;  
Best Local Similarity 66.7%; Pred. No. 3.4e-05;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDNDPOYKND 21  
DB 1 ANNHVSIVEAWSNDNDPTPLHD 21

RESULT 13  
US-09-604-957-5  
Sequence 5, Application US/09604957  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIKHUIZEN, LOBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604.957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 523  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Query Match 58.6%; Score 68; DB 6; Length 523;  
Best Local Similarity 66.7%; Pred. No. 0.012;  
Matches 14; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDNDPOYKND 21  
DB 75 ANOHLSTLEDSHNDPLYVTD 95

RESULT 14  
US-09-562-328-21  
Sequence 21, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562.328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 21  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-21

Query Match 56.9%; Score 66; DB 6; Length 21;  
Best Local Similarity 66.7%; Pred. No. 0.00061;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDNDPOYKND 21  
DB 1 ANDHSLIEAWSNDNDPTPLHD 21

RESULT 15  
US-09-604-957-6  
Sequence 6, Application US/09604957  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIKHUIZEN, LOBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604.957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 584  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-6

Query Match 55.2%; Score 64; DB 6; Length 584;  
Best Local Similarity 66.7%; Pred. No. 0.057;  
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDNDPOY 18  
DB 75 ANKHSLIEDWNGKRDPOY 92

Search completed: March 27, 2002, 14:22:47  
Job time: 1694 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:18 ; Search time 102.51 Seconds  
(without alignments)  
15.605 Million cell updates/sec

Title: US-09-290-049a-11  
Perfect score: 116  
Sequence: 1 AINHL5LEAWSNDNDPOYKND 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	1431	2 A45866	dextranucrase (EC
2	90	77.6	1577	2 T30858	glucosyltransferas
3	85	73.3	1449	2 T30857	glucosyltransferas
4	85	73.3	1449	2 T30552	glucosyltransferas
5	81	69.8	1365	2 A41483	glucosyltransferas
6	78	67.2	1508	2 T31098	probable dextranu
7	75	64.7	1475	2 B33135	glfB protein precu
8	74	63.8	1592	2 A38175	glucosyltransferas
9	68	58.6	1518	2 A44811	glucosyltransferas
10	67	57.8	1599	2 S22737	glucosyltransferas
11	66	56.9	1375	2 JT0345	dextranucrase (EC
12	49	42.2	632	2 T46504	hypothetical prote
13	47	40.5	175	2 C86205	hypothetical prote
14	47	40.5	175	2 B70961	probable esterase
15	47	40.5	403	2 FOXRL2	sigma 2 protein -
16	47	40.5	445	2 H75360	cytochrome P450 -
17	47	40.5	491	2 A86824	sensor protein kin
18	46	39.7	418	1 FOXRJS	sigma 2 protein -
19	46	39.7	420	2 T05877	hypothetical prote
20	46	39.7	623	2 T35377	probable membrane
21	46	39.7	626	2 D70178	PTS system, fructo
22	46	39.7	703	2 A48764	calpain (EC 3.4.22
23	45.5	39.2	250	2 G72495	probable polysulf
24	45	38.8	194	2 H72037	conserved hypobth
25	45	38.8	194	2 C86586	hypothetical prote
26	45	38.8	344	2 T21604	hypothetical prote
27	45	38.8	440	2 C84265	adenylosuccinate s
28	45	38.8	504	2 S51942	punin 2 precursor
29	45	38.8	637	2 T00548	hypothetical prote

30	45	38.8	657	2 S25184	cspl protein - Cor
31	45	38.8	1475	2 T29809	hypothetical prote
32	45	38.8	2358	2 T39569	probable alpha-glu
33	45	38.8	2371	2 T43432	alpha glucan synth
34	44.5	38.4	188	2 A75362	hypothetical prote
35	44.5	38.4	236	2 S54428	fimbrial protein h
36	44.5	38.4	241	2 S24978	fimbrial protein h
37	44	37.9	312	2 S67667	hypothetical prote
38	44	37.9	335	2 C85642	hypothetical prote
39	44	37.9	504	2 T24818	hypothetical prote
40	44	37.9	837	2 S54624	hypothetical prote
41	43.5	37.5	825	1 A60386	ROD1 protein - yea
42	43	37.1	267	1 S71020	interleukin-4 rece
43	43	37.1	267	2 C81937	peptidoglycan-link
44	43	37.1	267	2 H81167	competence lipopro
45	43	37.1	376	2 H86878	aminotransferase l

## ALIGNMENTS

```
RESULT 1
A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: A45866
R:Honda, O.; Kato, C.; Kuramitsu, H.K.
U. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans glfD gene encoding the gluco
A:Reference number: A45866; MUID:91100958
A:Accession: A45866
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: GB:M29296
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP6>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1365-1404/Domain: cpl repeat homology <CP7>

Query Match 100.0%; Score 116; DB 2; Length 1431;
Best local similarity 100.0%; Pred. No. 1.7e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AINHL5LEAWSNDNDPOYKND 21
Db 495 AINHL5LEAWSNDNDPOYKND 515

RESULT 2
T30858
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30858
R:Stimpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for
A:Reference number: 220909; MUID:95122197
A:Accession: T30858
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1577 <SIM>
A:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AA041413.1
C:Genetics:
```



A:Reference number: A33128  
A:Accession: A33128  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-171,173-641,'N',643-1475 <SH2>  
A:Experimental source: strain GS-5  
C:Superfamily: cpl repeat homology  
F:1096-1115/Domain: cpl repeat homology <CP1>  
F:1124-1243/Domain: cpl repeat homology <CP2>  
F:1289-1308/Domain: cpl repeat homology <CP3>  
F:1354-1373/Domain: cpl repeat homology <CP4>  
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 64.7%; Score 75; DB 2; Length 1475;  
Best Local Similarity 71.4%; Pred. No. 0.0035;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

1 AINHSTLEAWSNDPQYND 21  
|:|||||  
481 ANHSTLEAWSNDPQYND 501

RESULT 8  
A38175  
glucosyltransferase precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C:Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
C:Accession: A38175  
R:Abu, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus  
A:Reference number: A38175; MUID:91123227  
A:Accession: A38175  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1592 <RBO>  
A:Cross-references: GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID:d1014946; PID:g217033  
C:Superfamily: cpl repeat homology  
F:1093-1112/Domain: cpl repeat homology <CP1>  
F:1222-1241/Domain: cpl repeat homology <CP2>  
F:1287-1306/Domain: cpl repeat homology <CP3>  
F:1330-1351/Domain: cpl repeat homology <CP4>  
F:1352-1371/Domain: cpl repeat homology <CP5>  
F:1402-1420/Domain: cpl repeat homology <CP6>  
F:1465-1484/Domain: cpl repeat homology <CP7>  
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 63.8%; Score 74; DB 2; Length 1592;  
Best Local Similarity 66.7%; Pred. No. 0.0054;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

1 AINHSTLEAWSNDPQYND 21  
|:|||||  
477 ANHSTLEAWSNDPQYND 497

RESULT 9  
A44811  
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999  
C:Accession: A44811; S22726; S28809  
R:Gilford, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
A:Reference number: A44811; MUID:92148377  
A:Accession: A44811  
A:Molecule type: DNA  
A:Residues: 1-1518 <GIF>  
A:Cross-references: EMBL:Z11873; NID:g47526; PIDN:CAAT77900.1; PID:g47527  
A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIR:81052)

C:Genetics:  
A:Gene: gltf  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 58.6%; Score 68; DB 2; Length 1518;  
Best Local Similarity 63.2%; Pred. No. 0.043;  
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

1 AINHSTLEAWSNDPQYND 19  
|:|||||  
501 ALAHSTLEAWSNDPQYND 519

RESULT 10  
S22737  
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: S22737; S28810; B44811; S22727  
R:Jacques, N.  
Submitted to the EMBL Data Library, March 1992  
A:Reference number: S22726  
A:Accession: S22737  
A:Molecule type: DNA  
A:Residues: 1-1599 <JAC>  
A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAAT77898.1; PID:g47531  
R:Gilford, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
A:Reference number: A44811; MUID:92148377  
A:Accession: S28810  
A:Molecule type: DNA  
A:Residues: 1-51 <GIF>  
A:Cross-references: EMBL:Z11873  
C:Genetics:  
A:Gene: gltf  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1456-1475/Domain: cpl repeat homology <CP4>

Query Match 57.8%; Score 67; DB 2; Length 1599;  
Best Local Similarity 66.7%; Pred. No. 0.064;  
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

1 AINHSTLEAWSNDPQYND 18  
|:|||||  
491 ALAHSTLEAWSNDPQYND 508

RESULT 11  
JT0345  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
N:Alternate names: sucrose 6-glucosyltransferase  
C:Species: Streptococcus mutans  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
C:Accession: JT0345; C33135  
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
Gene 69, 101-109, 1988  
A:Title: Sequence analysis of the glfC gene from Streptococcus mutans GS-5.  
A:Reference number: JT0345; MUID:89137980  
A:Accession: JT0345  
A:Molecule type: DNA  
A:Residues: 1-1375 <UED>  
A:Experimental source: GS-5  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A:Title: Sequence analysis of the glfB gene from Streptococcus mutans.  
A:Reference number: A33135; MUID:87308013

A:Accession: C33135  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-349 <SHI>  
A:Cross-references: GB:M17361  
C:Genetics:  
A:Gene: gtfC  
C:Function:  
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
C:Superfamily: cpl repeat homology  
C:Keywords: duplication; glycosyltransferase; hexosyltransferase  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:35-1375/Product: glucosyltransferase #status predicted <MAT>  
F:1126-1145/Domain: cpl repeat homology <CP1>  
F:1253-1272/Domain: cpl repeat homology <CP2>  
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 56.9%; Score 66; DB 2; Length 1375;  
Best Local Similarity 66.7%; Pred. No. 0.077;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 AINHLSTLEAMSDNDPOYKND 21  
1 :||||| 1 1 1  
Db 507 ANDHLSTLEAMSYNDTPYLHD 527

RESULT 12  
T46504  
hypothetical protein DKFZp434M179.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T46504  
R:Postka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23029  
A:Accession: T46504  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-632 <AAA>  
A:Cross-references: EMBL:AL137338  
A:Experimental source: adult testis; clone DKFZp434M179  
C:Genetics:  
A>Note: DKFZp434M179.1

Query Match 42.2%; Score 49; DB 2; Length 632;  
Best Local Similarity 52.9%; Pred. No. 13;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 5 ISILEAMSDNDPOYKND 21  
5 :||| 1 :|||||  
Db 119 INVLAGASEFDPOYKND 135

RESULT 13  
C86205  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C86205  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chih, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hultar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719  
A:Accession: C86205

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-175 <STO>  
A:Cross-references: GB:AE005172; NID:98954041; PIDN:AAF82215.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 40.5%; Score 47; DB 2; Length 175;  
Best Local Similarity 46.2%; Pred. No. 6.3;  
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 2 INHSLTLEAMSDN 14  
2 :||| 1 :|||||  
Db 108 LNHQEVIAWSDH 120

RESULT 14  
B70961  
probable esterase - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: B70961  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon ; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987.  
A:Accession: B70961  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-403 <COL>  
A:Cross-references: GB:292669; GB:AL133456; NID:93242271; PIDN:CAB07015.1; PID:918715  
A:Experimental source: Strain H37RV  
C:Genetics:  
A:Gene: lipC  
C:Superfamily: Mycobacterium tuberculosis probable esterase

Query Match 40.5%; Score 47; DB 2; Length 403;  
Best Local Similarity 50.0%; Pred. No. 16;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 4 HSLTLEAMSDNDPOYKND 21  
4 :||| 1 :|||||  
Db 241 HSLALGILTANDPOYQAE 258

RESULT 15  
FOXRL2  
sigma 2 protein - reovirus type 1 (strain Lang)  
N:Alternate names: core protein  
C:Species: reovirus type 1  
A>Note: host Homo sapiens (man)  
C:Date: 31-Mar-1989 #sequence\_revision 30-Jun-1993 #text\_change 16-Jul-1999  
C:Accession: M1306; A29708  
R:Dermod, T.S.; Schiff, L.A.; Nibert, M.L.; Coombs, K.M.; Fields, B.N.  
J. Virol. 65, 5721-5731, 1991  
A:Title: The 52 gene nucleotide sequences of prototype strains of the three reovirus  
A:Reference number: M1306; MUID:92015462  
A:Accession: M1306  
A:Molecule type: genomic RNA  
A:Residues: 1-418 <DER>  
A:Cross-references: GB:S59098  
R:George, C.X.; Crowe, A.; Munemitsu, S.M.; Atwater, J.A.; Samuel, C.E.  
Biochem. Biophys. Res. Commun. 147, 1153-1161, 1987  
A:Title: Biosynthesis of reovirus-specified polypeptides. Molecular cDNA cloning and  
-2.  
A:Reference number: A29708; MUID:88024195  
A:Accession: A29708  
A:Molecule type: genomic RNA



A:Residues: 1-88,'V',90-204,'Y',206-322,'VATSCMWSKCDEW' <GEO>  
A:Cross-references: GB:MI7598; NID:g333748; PIDN:AAA47278.1; PID:g333749  
C:Genetics:  
A:Map position: segment S2  
C:Superfamily: reovirus sigma 2 protein  
C:Keywords: core protein

Query Match	40:58;	Score 47;	DB 1;	Length 418;
Best Local Similarity	61.58;	Pred. NO. 17;		
Matches	8;	Conservative	2;	Mismatches
			3;	Indels
				0;
				Gaps
				0;
OY	8	LEAWSDNDPOYNK	20	
		:      :		
Db	372	LEAWAREDDOYNQ	384	

Search completed: March 27, 2002, 14:01:19  
Job time: 481 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:57:54 ; Search time 198.55 Seconds  
(without alignments)  
7.834 Million cell updates/sec

Title: US-09-290-049a-10

Perfect score: 119  
Sequence: 1 ANDHLSTLEAMSDNDPTLYLHD 21

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_1101.\*  
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	90.8	1592	14	AA1932925
2	63	52.9	1577	17	AA1930417
3	60	50.4	2057	21	AA1910667
4	51	42.9	486	22	AA1936552
5	50	42.0	305	21	AA1909845
6	50	42.0	339	21	AA1909844
7	48.5	40.8	401	22	AA1935444
8	47.5	39.9	534	21	AA197414
9	47	38.5	195	21	AA1909941
10	47	38.5	338	15	AA195965
11	46	38.7	118	20	AA1960021

12	46	38.7	199	22	AA190817
13	46	38.7	336	20	AA193244
14	45	37.8	236	20	AA193993
15	45	37.8	385	12	AA19238
16	44	37.0	62	21	AA194863
17	44	37.0	161	21	AA1928444
18	44	37.0	206	21	AA1928443
19	44	37.0	220	22	AA1928161
20	44	37.0	384	21	AA1928041
21	43	36.1	126	21	AA1928416
22	43	36.1	126	21	AA193894
23	43	36.1	170	21	AA1926415
24	43	36.1	170	21	AA193893
25	43	36.1	193	21	AA193892
26	43	36.1	194	21	AA1928414
27	43	36.1	327	22	AA196817
28	43	36.1	335	20	AA1907036
29	43	36.1	335	20	AA1905368
30	43	36.1	1195	20	AA1975420
31	43	36.1	2466	16	AA1971498
32	43	36.1	2466	19	AA1975999
33	43	36.1	2466	21	AA190272
34	43	36.1	2485	21	AA19343
35	42.5	35.7	325	22	AA1941942
36	42.5	35.7	390	22	AA1940156
37	42	35.3	123	22	AA197865
38	42	35.3	174	21	AA1960587
39	42	35.3	178	10	AA190061
40	42	35.3	185	21	AA1960586
41	42	35.3	189	21	AA1960585
42	42	35.3	236	20	AA1960021
43	42	35.3	239	20	AA198935
44	42	35.3	259	17	AA1989326
45	42	35.3	298	19	AA1971486

#### ALIGNMENTS

RESULT 1	AA1932925	standard; Protein; 1592 AA.
ID	AA1932925	
XX		
AC	AA1932925	
XX		
DT	28-JUN-1993	(first entry)
XX		
DE	Glucosyltransferase I.	
XX		
KW	GT-1; Streptococcus; dental; caries.	
OS	Streptococcus sobrinus.	
XX		
PN	JP05023188-A.	
XX		
DR	02-FEB-1993.	
PD		
XX		
PF	25-JUL-1991;	91JP-0186592.
XX		
PR	25-JUL-1991;	91JP-0186592.
XX		
PA	(FUKU/) FUKU I.	
XX		
PA	(KATO/) KATO K.	
XX		
DR	WPI: 1993-079449/10.	
XX		
PT	N-PSDB: AAQ37760.	
XX		
PT	DNA sequence glucosyl:transferase-I - comprises Streptococcus	
XX	sobrinus DNA sequence with at least one nucleotide added or	
XX	deleted	
PS	Claim 13; Page 15; 29pp; Japanese.	

CC The DNA sequence from *Streptococcus sobrinus* strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating  
 CC *S. sobrinus* 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with SauAI and fractionating on agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and *E. coli* JM109  
 CC transformed with it. A GT-1 expressing clone was isolated and  
 CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 CC  
 XX Sequence 1592 AA;

Query Match 90.8%; Score 108; DB 14; Length 1592;  
 Best Local Similarity 85.7%; Pred. No. 1.2e-07;  
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANHLSLEAMSDNDPYLHD 21  
 |||:|||||  
 477 anhvslveawsdndpylhd 497

## RESULT 2

AAR91047  
 ID AAR91047 standard; Protein; 1577 AA.

AC AAR91047;

DT 22-MAY-1996 (first entry)

DE Alpha-D-glucosyltransferase.

XX Alpha-D-glucosyltransferase; primer-independent; soluble glucan;  
 KM sucrose; transgenic plant; cloning; *Escherichia coli*;  
 KM phage lambda-CI3; vector; plasmid pGSG501; plasmid pGSG502;  
 KM gene transfer; crop improvement; storage carbohydrate; pasture;  
 KM feedstuff; senescence; dextran; binder; food; pharmaceutical.  
 XX  
 OS *Streptococcus salivarius* strain ATCC 25975.

XX WO9606173-A1.

XX 29-FEB-1996.

XX 24-AUG-1995; 95WO-AU00527.

XX 24-AUG-1994; 94AU-0007643.

PA (GIEF/) GIFFARD P M.  
 XX (JACO/) JACQUES N A.  
 XX (SIMP/) SIMPSON C L.

PI Giffard PM, Jacques NA, Simpson CL;

XX WPI; 1996-151376/15.

DR N-PSDB; AAT13139.

XX  
 PT Plants cont. new bacterial DNA encoding glucosyl transferase  
 PT activity - retain higher levels of stored carbohydrate(s) in a form  
 PT readily digestible by ruminants

XX Claim 4: Page 16-20; 31pp; English.

XX The sequence represents an alpha-D-glucosyltransferase from  
 CC *Streptococcus salivarius*. The enzyme is primer-independent, and  
 CC produces soluble glucan from sucrose. A gene encoding the enzyme  
 CC may be cloned and expressed in *Escherichia coli* using a subclone  
 CC of phage lambda-CI3, e.g. plasmid pGSG501 or plasmid pGSG502. The  
 CC DNA may also be expressed in a transgenic plant, to improve the  
 CC level of stored carbohydrate in a pasture plant which normally  
 CC contains low levels, or to prevent degradation of stored carbohydrate  
 CC during plant senescence. Dextran may be isolated from the plant, for  
 CC use as a food binder or pharmaceutical additive. Primer independence  
 CC ensures that the enzyme will be functional in plants. The glucan is

CC poorly degraded in plants but easily degraded by bacteria in the rumen  
 CC of grazing livestock.  
 CC  
 XX Sequence 1577 AA;

Query Match 52.9%; Score 63; DB 17; Length 1577;  
 Best Local Similarity 72.2%; Pred. No. 0.74;  
 Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 HLSLEAMSDNDPYLHD 21  
 |||:|||||  
 Db 591 hlsleawsyndpylhd 608

## RESULT 3

AAB10667  
 ID AAB10667 standard; Protein; 2057 AA.

AC AAB10667;

DT 19-JAN-2001 (first entry)

DE L. mesenteroides alternan sucrose protein.

XX Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 KM syrup.

XX Leuconostoc mesenteroides.

XX DE19905069-A1.

XX 10-AUG-2000.

XX 08-FEB-1999; 99DE-1005069.

XX 08-FEB-1999; 99DE-1005069.

PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
 XX (PLAC) MAX PLANCK GES FÖRDERUNG WISSENSCHAFTEN.

XX Kossman J, Welsh T, Quanz M, Knuth K;

XX WPI; 2000-550294/51.

DR N-PSDB; AAA97904.

XX New nucleic acid encoding recombinant Leuconostoc mesenteroides  
 PT alternan sucrose protein and methods of alternan and fructose  
 PT production

XX Claim 1a: Page 30-36; 64pp; German.

XX This invention describes a novel nucleic acid molecule (I) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group) The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The alternan and/or  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC which previously has been achieved by costly production from maize  
 CC starch. This sequence represents the leuconostoc mesenteroides alternan  
 CC sucrose protein which is described in the method of the invention.

XX Sequence 2057 AA;

Query Match 50.4%; Score 60; DB 21; Length 2057;  
 Best Local Similarity 55.0%; Pred. No. 2.8;  
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Oy 1 ANDHLSLEAMSDNDTPYLH 20  
||| ||||| | : :  
Db 665 ankhsliedwngkdpqyv 684

RESULT 4

AAB36552  
ID AAB36552 standard; Protein: 486 AA.

AC AAB36552;

DT 07-MAR-2001 (first entry)

DE Lawsonia intracellularis flagellar hook protein FlgE SEQ ID NO:1.

KW Lawsonia intracellularis; flgE; flagellar hook protein; vaccine;  
KW intestinal disease; immunogenic; diagnosis; antibacterial; swine;  
KW pig; infection; detection; identification.

XX Lawsonia intracellularis.

WO200069904-A1.

PD 23-NOV-2000.

PF 11-MAY-2000; 2000WO-AU00437.

PR 13-MAY-1999; 99US-0133973.

PA (PF12 ) PFIZER PROD INC.  
(AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
(PIGR-) PIG RES & DEV CORP.

PI Panaccio M, Rosey EL, Sinistaj M, Hasse D, Parsons J;  
Ankenbauer RG;

PI Ankenbauer RG;

DR WPI: 2001-016210/02.

DR N-PSDB: AAC88037.

XX New immunogenic Lawsonia FlgE peptide, its nucleic acid and antibody,  
PT useful in vaccines and diagnosis of Lawsonia infections, particularly  
PT in swine -

PS Claim 13; Page 87-90; 97pp; English.

XX The present sequence is the Lawsonia intracellularis flagellar hook  
CC protein FlgE. The present invention describes an isolated or recombinant  
CC polypeptide (I) that comprises, mimics or cross-reacts with a B- or  
CC T-cell epitope of a FlgE (flagellar hook) polypeptide from a

CC Lawsonia spp. (I) has antibacterial activity, and induces a specific  
CC humoral immune response. (I) are used as antigens in vaccines to prevent  
CC or treat infection by Lawsonia, in birds and animals, especially pigs,  
CC to raise specific antibodies (Ab) and to detect past or present  
CC infection. Ab are also useful in diagnosis, to detect L. intracellularis  
CC or immunologically cross-reactive species, also for identification of  
CC epitopes in FlgE. Vectors that contain nucleic acids (II) encoding (I)  
CC are also useful in genetic vaccines, and fragments of (II) are useful  
CC as primers or probes for detecting L. intracellularis or related  
CC microorganisms, in hybridisation or amplification assays.

SO Sequence 486 AA;

Query Match 42.9%; Score 51; DB 22; Length 486;  
Best Local Similarity 47.1%; Pred. NO. 12;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Oy 1 ANDHLSLEAMSDNDTP 17  
|| : : || : | | |

Db 169 anpyrtalleswngntp 185

RESULT 5

AAG09845  
ID AAG09845 standard; Protein: 305 AA.

AC AAG09845;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 7935.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 28-APR-1999; 99US-0130510.

PR 30-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 04-MAY-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 17-JUN-1999; 99US-0139453.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR	18-JUN-1999	9905-01394559
PR	18-JUN-1999	9905-01394660
PR	18-JUN-1999	9905-01394661
PR	18-JUN-1999	9905-01394662
PR	18-JUN-1999	9905-01394663
PR	18-JUN-1999	9905-01397501
PR	18-JUN-1999	9905-01397663
PR	21-JUN-1999	9905-01398617
PR	22-JUN-1999	9905-01398699
PR	23-JUN-1999	9905-01400353
PR	23-JUN-1999	9905-01400354
PR	23-JUN-1999	9905-01400355
PR	24-JUN-1999	9905-01400356
PR	28-JUN-1999	9905-01400823
PR	29-JUN-1999	9905-01400891
PR	30-JUN-1999	9905-01411287
PR	01-JUL-1999	9905-01418542
PR	01-JUL-1999	9905-01421551
PR	02-JUL-1999	9905-01420055
PR	06-JUL-1999	9905-01423909
PR	08-JUL-1999	9905-01428033
PR	09-JUL-1999	9905-01429290
PR	12-JUL-1999	9905-01429277
PR	13-JUL-1999	9905-01435442
PR	14-JUL-1999	9905-01433624
PR	15-JUL-1999	9905-01440085
PR	16-JUL-1999	9905-01440085
PR	16-JUL-1999	9905-01440086
PR	19-JUL-1999	9905-01444335
PR	19-JUL-1999	9905-01444335
PR	19-JUL-1999	9905-01444335
PR	19-JUL-1999	9905-01444331
PR	19-JUL-1999	9905-01444332
PR	19-JUL-1999	9905-01444333
PR	19-JUL-1999	9905-01443864
PR	21-JUL-1999	9905-01448814
PR	21-JUL-1999	9905-01450086
PR	21-JUL-1999	9905-01450088
PR	22-JUL-1999	9905-01450085
PR	22-JUL-1999	9905-01450087
PR	22-JUL-1999	9905-01450089
PR	22-JUL-1999	9905-01451592
PR	23-JUL-1999	9905-01451592
PR	23-JUL-1999	9905-01452118
PR	23-JUL-1999	9905-01455224
PR	26-JUL-1999	9905-01455276
PR	27-JUL-1999	9905-01455913
PR	27-JUL-1999	9905-01455916
PR	28-JUL-1999	9905-01459192
PR	02-AUG-1999	9905-01463866
PR	02-AUG-1999	9905-01463868
PR	03-AUG-1999	9905-01470389
PR	04-AUG-1999	9905-01472034
PR	04-AUG-1999	9905-01473092
PR	05-AUG-1999	9905-01473533
PR	05-AUG-1999	9905-01473533
PR	06-AUG-1999	9905-01473603
PR	06-AUG-1999	9905-01474166
PR	09-AUG-1999	9905-01474935
PR	09-AUG-1999	9905-01474935
PR	16-AUG-1999	9905-01493684
PR	17-AUG-1999	9905-01491755
PR	18-AUG-1999	9905-01494265
PR	20-AUG-1999	9905-01497222
PR	20-AUG-1999	9905-01497223

PR	20-AUG-1999;	9905-0149929;
PR	23-AUG-1999;	9905-0149929;
PR	23-AUG-1999;	9905-0149930;
PR	26-AUG-1999;	9905-0150084;
PR	27-AUG-1999;	9905-0151065;
PR	27-AUG-1999;	9905-0151066;
PR	27-AUG-1999;	9905-0151080;
PR	30-AUG-1999;	9905-0151033;
PR	31-AUG-1999;	9905-0151438;
PR	01-SEP-1999;	9905-0151930;
PR	07-SEP-1999;	9905-0152363;
PR	10-SEP-1999;	9905-0153070;
PR	13-SEP-1999;	9905-0153578;
PR	15-SEP-1999;	9905-0154039;
PR	16-SEP-1999;	9905-0154018;
PR	20-SEP-1999;	9905-0155139;
PR	22-SEP-1999;	9905-0155439;
PR	23-SEP-1999;	9905-0153486;
PR	24-SEP-1999;	9905-0155659;
PR	28-SEP-1999;	9905-0156458;
PR	29-SEP-1999;	9905-0156596;
PR	04-OCT-1999;	9905-0157117;
PR	05-OCT-1999;	9905-0157753;
PR	06-OCT-1999;	9905-0157865;
PR	07-OCT-1999;	9905-0158029;
PR	08-OCT-1999;	9905-0158312;
PR	12-OCT-1999;	9905-0158369;
PR	13-OCT-1999;	9905-0159293;
PR	13-OCT-1999;	9905-0159294;
PR	14-OCT-1999;	9905-0159329;
PR	14-OCT-1999;	9905-0159330;
PR	14-OCT-1999;	9905-0159331;
PR	14-OCT-1999;	9905-0159637;
PR	14-OCT-1999;	9905-0159638;
PR	18-OCT-1999;	9905-0159584;
PR	21-OCT-1999;	9905-0160741;
PR	21-OCT-1999;	9905-0160767;
PR	21-OCT-1999;	9905-0160768;
PR	21-OCT-1999;	9905-0160710;
PR	21-OCT-1999;	9905-0160814;
PR	21-OCT-1999;	9905-0160815;
PR	22-OCT-1999;	9905-0160880;
PR	22-OCT-1999;	9905-0160981;
PR	22-OCT-1999;	9905-0160989;
PR	25-OCT-1999;	9905-0161404;
PR	25-OCT-1999;	9905-0161405;
PR	25-OCT-1999;	9905-0161406;
PR	26-OCT-1999;	9905-0161359;
PR	26-OCT-1999;	9905-0161360;
PR	26-OCT-1999;	9905-0161361;
PR	28-OCT-1999;	9905-0161920;
PR	28-OCT-1999;	9905-0161982;
PR	28-OCT-1999;	9905-0161993;
PR	29-OCT-1999;	9905-0162142;

Query Match	42.0%;	Score 50;	DB 21;	Length 305;
Best Local Similarity	52.9%;	Pred. No. 10;		
Matches	9;	Conservative	3;	Mismatches
			5;	Indels
				0;
Gaps				0;
Oy	3	DHLSTLEAMSDNDPPYL	19	
	:	:	:	
Db	200	dyeqimeawsdkgilyv	216	
RESULT	6			
AAAG09844				
XX	AAAG09844	standard;	Protein;	339 AA.
XX				
AC	AAAG09844;			
XX				
DT	17-OCT-2000	(first entry)		

XX Arabidopsis thaliana protein fragment SEQ ID NO: 7934.  
DE XX  
KW Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
OS Arabidopsis thaliana.  
PN EPI033405-A2.  
XX  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 03-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142927.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145112.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147132.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.

PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	42.0%;	Score 50;	DB 21;	Length 339;
Best Local Similarity	52.9%;	Pred. NO. 12;		
Matches	9;	Conservative	3;	Mismatches
			5;	Indels
			0;	Gaps
Oy	3	DHLSLEAMSDNDPPYL	19	
	: :	: :		
Db	234	dyeimeawsdskgtlyv	250	
RESULT	7			
AAAM43544				
ID	AAAM43544	standard;	Protein:	401 AA.
XX				
AC	AAAM43544;			
XX				
DT	22-OCT-2001	(first entry)		
XX				
DE	Human polypeptide SEQ	ID NO 222.		
XX				
KM	Human: antiarthritic; antirheumatic; antiproliferative; vasotropic;			
KM	cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;			
KM	fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;			
KM	neuroprotective; antiallergic; hepatotropic; antidiabetic;			

XW		antiinflammatory; antituber; vulnereary; anticonvulsant; antibacterial;
KW		antiparasitic; cardiant; gene therapy; cancer; immune disorder;
KW		cardiovascular disorder; neurological disease; infection; human.
XX		
OS	Homo sapiens.	
XX		
PN	WO200155308-A2.	
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001; 2001WO-US01309.	
XX		
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209467.	
PR	28-JUN-2000; 2000US-0214886.	
PR	30-JUN-2000; 2000US-0215135.	
PR	07-JUL-2000; 2000US-0216547.	
PR	07-JUL-2000; 2000US-0216880.	
PR	11-JUL-2000; 2000US-0217487.	
PR	11-JUL-2000; 2000US-0217496.	
PR	14-JUL-2000; 2000US-0218290.	
PR	26-JUL-2000; 2000US-0220963.	
PR	26-JUL-2000; 2000US-0220964.	
PR	14-AUG-2000; 2000US-0224518.	
PR	14-AUG-2000; 2000US-0224519.	
PR	14-AUG-2000; 2000US-0225213.	
PR	14-AUG-2000; 2000US-0225214.	
PR	14-AUG-2000; 2000US-0225266.	
PR	14-AUG-2000; 2000US-0225267.	
PR	14-AUG-2000; 2000US-0225268.	
PR	14-AUG-2000; 2000US-0225270.	
PR	14-AUG-2000; 2000US-0225447.	
PR	14-AUG-2000; 2000US-0225757.	
PR	14-AUG-2000; 2000US-0225758.	
PR	14-AUG-2000; 2000US-0225759.	
PR	18-AUG-2000; 2000US-0226279.	
PR	22-AUG-2000; 2000US-0226681.	
PR	22-AUG-2000; 2000US-0226868.	
PR	22-AUG-2000; 2000US-0227182.	
PR	23-AUG-2000; 2000US-0227209.	
PR	30-AUG-2000; 2000US-0228924.	
PR	01-SEP-2000; 2000US-0229287.	
PR	01-SEP-2000; 2000US-0229343.	
PR	01-SEP-2000; 2000US-0229344.	
PR	01-SEP-2000; 2000US-0229345.	
PR	05-SEP-2000; 2000US-0229509.	
PR	05-SEP-2000; 2000US-0229513.	
PR	06-SEP-2000; 2000US-0230437.	
PR	06-SEP-2000; 2000US-0230438.	
PR	08-SEP-2000; 2000US-0231142.	
PR	08-SEP-2000; 2000US-0231143.	
PR	08-SEP-2000; 2000US-0231244.	
PR	08-SEP-2000; 2000US-0231244.	
PR	08-SEP-2000; 2000US-0231413.	
PR	08-SEP-2000; 2000US-0231414.	
PR	08-SEP-2000; 2000US-0232080.	
PR	08-SEP-2000; 2000US-0232081.	
PR	12-SEP-2000; 2000US-0231568.	
PR	14-SEP-2000; 2000US-0232297.	
PR	14-SEP-2000; 2000US-0232298.	
PR	14-SEP-2000; 2000US-0232399.	
PR	14-SEP-2000; 2000US-0232400.	
PR	14-SEP-2000; 2000US-0232401.	
PR	14-SEP-2000; 2000US-0233063.	
PR	14-SEP-2000; 2000US-0233064.	
PR	14-SEP-2000; 2000US-0233065.	
PR	21-SEP-2000; 2000US-0234423.	



PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0251989.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-488781/53.  
 DR N-PSDB; AA163850.  
 XX  
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,  
 PT treating and/or preventing human diseases and disorders -  
 XX  
 PS Claim 11: SEQ ID NO 222; 664bp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AA163803-AA164012) and  
 CC the encoded proteins (AA1634497-AA163660) useful for preventing, treating  
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The  
 CC genes were isolated from a range of human tissues disclosed in the  
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists  
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer;  
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,  
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 401 AA;  
 Query Match 40.8%; Score 48.5; DB 22; Length 401;  
 Best Local Similarity 47.4%; Pred. No. 24;  
 Matches 9; Conservative 3; Mismatches 4; Indels 3; Gaps 1;  
 QY 5 LSTLEAMSDN--DPYLAH 20  
 :::||| | | | | |  
 Db 46 vallevwkdntddxpxh 64  
 RESULT 8  
 ID: AA167414 standard; protein; 534 AA.  
 XX  
 AC AA167414;  
 XX  
 DT 12-MAY-2000 (first entry)  
 XX  
 DE Arabidopsis aldehyde dehydrogenase (ALDH)-3.  
 XX  
 KW Arabidopsis; plant plastid; acetyl CoA synthetase; E3 subunit; ACS;  
 KW plant plastidic pyruvate dehydrogenase; PPDH; ATP citrate lyase; ACL;  
 KW pyruvate decarboxylase; PDC; aldehyde dehydrogenase; ALDH; acetyl CoA;  
 KW fatty acid; flavonoid; enzyme; phytochemical; pyruvate decarboxylase;  
 KW acetyl CoA hydrolase; mitochondrial pyruvate dehydrogenase.  
 XX  
 OS Arabidopsis sp.  
 XX  
 PN WO200000619-A2.  
 PN  
 PD 06-JAN-2000.  
 XX  
 PE 25-JUN-1999; 99WO-US14382.  
 XX  
 PR 26-JUN-1998; 98US-0090717.  
 PR  
 PA (IOWA ) UNIV IOWA STATE RES FOUND INC.

XX Nikolau BJ, Murtele ES, Oliver DJ, Behal R, Schnable PS, Ke J;  
PI Johnson JL, Allied CC, Fatland B, Lutziger I, Wen T;  
XX  
DR WPI: 2000-160678/14.  
N-PSDB: AA556975.  
XX  
XX Novel acetyl CoA synthetase (ACS), plastidic pyruvate dehydrogenase  
PT (PPDH), Arp citrate lyase (ACL), pyruvate decarboxylase (PDC) and  
PT aldehyde dehydrogenase (ALDH) polypeptides used to alter acetyl CoA  
PT levels in plants  
XX  
XX Examples: Fig 12B; 79pp; English.  
XX  
CC The invention provides nucleic acids encoding Arabidopsis plastidic  
CC acetyl CoA synthetase (ACS), various subunits (specifically the E3  
CC subunit) of plant plastidic pyruvate dehydrogenase (PPDH), the A and B  
CC subunits of a plant Arp citrate lyase (ACL), Arabidopsis pyruvate  
CC decarboxylase (PDC), Arabidopsis aldehyde dehydrogenase (ALDH),  
CC specifically ALDH-2 and ALDH-4. The polypeptides can be expressed by  
CC standard recombinant methodology. The ACS, PPDH, ACL, PDC and ALDH  
CC polypeptides, methods and nucleic acid molecules of the invention are  
CC used to alter the level of acetyl CoA in a plant or plant cell, tissues  
CC or organs. A decrease in acetyl CoA is expected to affect the  
CC biosynthesis of very long chain fatty acids and flavonoids. The enzymes  
CC may also be used for in vitro synthesis of acetyl CoA, which in turn can  
CC be used to produce acetyl CoA phytochemicals (plastidic ACS, PPDH, ACL,  
CC pyruvate decarboxylase, acetyl CoA hydrolyase, mitochondrial pyruvate  
CC dehydrogenase and aldehyde dehydrogenase are involved in the biosynthesis  
CC of acetyl CoA.  
XX  
SQ Sequence 534 AA:

Query Match 39.9%; Score 47.5; DB 21; Length 534;  
Best Local Similarity 52.9%; Pred. No. 47;  
Matches 9; Conservative 2; Mismatches 5; Indels 1; Gaps 1;  
QY 2 NDHLSLEAMSDNTPY 18  
||::|||  
DB 132 ndelaaletw-dngkpy 147

RESULT 9  
AAG09941  
ID AAG09941 strand: Protein; 195 AA.

AAG09941;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 8067.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130049.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137502.  
PR 08-JUN-1999; 99US-0137724.  
PR 10-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0138119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 39.5%; Score 47; DB 21; Length 195;  
Best Local Similarity 36.8%; Pred. No. 17;  
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 3 DHLSLEAMSDMDTPYLHD 21  
Db 109 nhqewidawsdhqkplwtd 127

RESULT 10

AAR65965  
ID AAR65965 standard; Protein; 338 AA.

AC AAR65965;

DT 06-JUN-1995 (first entry)

XX T. niveum GAPDH.

XX Alanine-racemase; glyceraldehyde-3-phosphate-dehydrogenase;

KW GAPDH; cyclosporin; immunosuppressive.

XX Tolypocladium niveum ATCC 34921.

XX OS

XX W09425606-A.

XX PN

XX PD 10-NOV-1994.

XX PP 23-APR-1994; 94WO-EP01272.

XX PR 23-APR-1993; 93DE-4312856.

XX PR 04-MAY-1993; 93DE-4314610.

XX PR 17-MAY-1993; 93DE-4316419.

XX PA (SANO ) SANDOZ LTD.

XX PA (SANO ) SANDOZ PATENT GMBH.

XX PI (SANO ) SANDOZ-ERFINDUNGEN VERW GRS MBH.

XX PI Kocher HP, Schneider-scherzer E, Schoergendorfer K;

XX DR WPT; 1994-358281/44.

XX New nucleic acid encoding eukaryotic alanine racemase - and  
PT related vectors, host cells and recombinant enzyme, useful for  
PT producing cyclosporin derivs. or increasing cyclosporin prodn.,

PT also new glyceraldehyde-3-phosphate dehydrogenase gene.  
 XX  
 PS Disclosure; Fig.15; 82pp; German.  
 XX  
 CC A cosmid bank of T. niveum ATCC 34921 genomic DNA was screened with  
 CC a fragment of the glyceraldehyde-3-phosphate-dehydrogenase (GAPDH)  
 CC gene from *Penicillium chrysogenum*. A 2.2 kb. fragment was  
 CC identified (AA078277) and cloned in pUC18 to give pGT1. A 665 bp  
 CC HindIII-HindIII fragment of this was used to screen cosmid DNA to  
 CC identify a DNA fragment (AA078279) containing the GAPDH promoter.  
 CC A sequence (AA078278) encoding the GAPDH gene (gpdh) was isolated from  
 CC a cDNA library. The full sequence of the genomic gpdh structural  
 CC gene is given in AA078280 and the corresponding protein sequence in  
 CC AAR5965. The GAPDH promoter is useful for expressing foreign genes  
 CC in T. niveum.  
 XX  
 SQ Sequence 338 AA:  
 OY  
 DB 304 ndhfaklsvwydnewgy 320  
 OY 2 NDHLSLEAWSNDTPY 18  
 DB 304 ndhfaklsvwydnewgy 320

RESULT 11  
 AAY60021  
 ID AAY60021 standard; Protein; 118 AA.  
 AC AAY60021;  
 XX  
 DT 31-JAN-2000 (first entry)  
 XX  
 DE Human endometrium tumour EST encoded protein 81.  
 XX  
 KM Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:  
 KM treatment; uterine; gene therapy; expressed sequence tag.  
 XX  
 OS Homo sapiens.  
 XX  
 PN DE19817948-A1.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PR 17-APR-1998; 98DE-1017948.  
 XX  
 PR 17-APR-1998; 98DE-1017948.  
 XX  
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 XX  
 PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
 XX  
 DR WPI; 1999-591957/51.  
 XX  
 DR N-PSDB; AA242007.  
 XX  
 PT New nucleic acid sequences expressed in uterine cancer tissues, and  
 PT derived polypeptides, for treatment of uterine and endometrial cancer  
 PT and identification of therapeutic agents -  
 XX  
 PS Claim 23; Page 307; 444pp; German.  
 XX  
 CC This invention describes novel human nucleic acid (cDNA) sequences (A),  
 CC that are highly expressed in uterine tumour tissue and which have  
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
 CC are used (i) to identify agents suitable for treatment of uterine or  
 CC endometrial cancer; (ii) directly for treating these forms of cancer  
 CC (including expression from gene therapy vectors) and (iii) for  
 CC generation of specific antibodies. (A) are identified by assembling ESTs  
 CC (expressed sequence tags) from a particular tissue type before comparison

CC of expression patterns. This allows a significantly longer fragment of  
 CC the gene to be revealed, so should reduce the number of failures  
 CC associated with the fact that ESTs from different libraries may represent  
 CC different parts of the same unknown gene, distorting the estimated  
 CC frequency of occurrence in a particular tissue. AAY5941-Y60328 represent  
 CC protein fragments encoded by the human endometrium tumour cDNA library  
 CC derived EST fragments represented in AA241981-242121.  
 XX  
 SQ Sequence 118 AA:

Query Match 38.7%; Score 46; DB 20; Length 118;  
 Best Local Similarity 50.0%; Pred. No. 14;  
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
 OY 3 DHLSLEAWSNDTPYLH 20  
 DB 23 dplldldagqdkdtpssh 40

RESULT 12  
 AAG90817  
 ID AAG90817 standard; Protein; 199 AA.  
 AC AAG90817;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum protein fragment SEQ ID NO: 4571.  
 XX  
 KM Corynebacterium; amino acid synthesis; vitamin; saccharide;  
 KM organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EP1108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 PA (KYO) KYOWA HAKKO KOGYO KK.  
 XX  
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 DR WPI; 2001-376931/40.  
 DR N-PSDB; AAH60036.  
 XX  
 PT Novel polynucleotides derived from Corynebacterium, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 PS Claim 17; SEQ ID NO: 4571; 246pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Corynebacterium glutamicum glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of corynebacterium glutamicum, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Corynebacterium glutamicum, and identifying a homologue of a gene derived  
 CC from corynebacterium glutamicum. Corynebacterium are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein : protein search, using sw model

Run on: March 27, 2002, 13:59:28 ; Search time 87.3 Seconds  
(without alignments)  
5.413 Million cell updates/sec

Title: US-09-290-049A-10

Perfect score: 119  
Sequence: 1 ANDHLSILEAWSNDNIPYLDH 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents-AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	1475	US-09-007-999-2	Sequence 2, Appli
2	119	100.0	1475	US-09-210-361-2	Sequence 2, Appli
3	110	92.4	1375	US-09-210-361-4	Sequence 4, Appli
4	75	63.0	1430	US-09-008-172-2	Sequence 2, Appli
5	75	63.0	1430	US-09-210-361-6	Sequence 6, Appli
6	63	52.9	1577	US-08-793-824-2	Sequence 2, Appli
7	45	37.8	385	US-08-361-920-23	Sequence 23, Appli
8	45	37.8	385	US-08-479-939-23	Sequence 23, Appli
9	45	37.8	385	US-08-483-432-23	Sequence 23, Appli
10	44	37.0	384	US-09-311-170-2	Sequence 2, Appli
11	43	36.1	2465	US-08-596-291-3	Sequence 3, Appli
12	43	36.1	2465	US-09-100-804-3	Sequence 3, Appli
13	43	36.1	2466	US-09-080-855-12	Sequence 12, Appli
14	43	36.1	2466	PCT-US94-09943-2	Sequence 2, Appli
15	43	36.1	2485	US-09-290-640-46	Sequence 46, Appli
16	42	35.3	259	US-08-277-231A-3	Sequence 3, Appli
17	42	35.3	259	US-08-473-750-6	Sequence 6, Appli
18	42	35.3	259	US-08-477-326-6	Sequence 6, Appli
19	42	35.3	259	US-08-070-301-16	Sequence 16, Appli
20	41	34.5	480	US-08-272-255-12	Sequence 12, Appli
21	41	34.5	480	PCT-US95-08565-12	Sequence 12, Appli
22	40	33.6	196	US-08-933-750C-35	Sequence 35, Appli
23	40	33.6	196	US-09-234-613-35	Sequence 35, Appli
24	40	33.6	307	US-08-332-562A-83	Sequence 83, Appli
25	40	33.6	334	US-09-362-473-6	Sequence 6, Appli
26	40	33.6	455	US-09-362-473-6	Sequence 6, Appli
27	40	33.6	761	US-09-625-188-14	Sequence 14, Appli

28	40	33.6	839	4	US-09-197-636-2	Sequence 2, Appli
29	40	33.6	839	4	US-09-197-636-4	Sequence 4, Appli
30	40	33.6	839	4	US-09-197-636-8	Sequence 8, Appli
31	40	33.6	2860	2	US-08-826-267-2	Sequence 2, Appli
32	40	33.6	3898	4	US-08-750-717-2	Sequence 2, Appli
33	39.5	33.2	337	3	US-09-032-372-2	Sequence 3, Appli
34	39.5	33.2	500	6	5171684-2	Patent No. 5171684
35	39.5	33.2	567	1	US-08-258-261B-4	Sequence 4, Appli
36	39.5	33.2	567	1	US-08-456-837-4	Sequence 4, Appli
37	39.5	33.2	567	1	US-08-457-342-4	Sequence 4, Appli
38	39.5	33.2	567	1	US-08-457-646A-4	Sequence 4, Appli
39	39.5	33.2	567	1	US-08-458-076A-4	Sequence 4, Appli
40	39.5	33.2	567	1	US-08-457-335A-4	Sequence 4, Appli
41	39.5	33.2	567	2	US-08-729-214-4	Sequence 4, Appli
42	39.5	33.2	567	2	US-09-028-934-4	Sequence 4, Appli
43	39	32.8	755	3	US-09-071-101-2	Sequence 4, Appli
44	39	32.8	755	3	US-09-369-618-2	Sequence 2, Appli
45	39	32.8	755	3	US-09-369-617-2	Sequence 2, Appli

#### ALIGNMENTS

```

RESULT 1
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match 100.0%; Score 119; DB 3; Length 1475;
Best Local Similarity 100.0%; Pred. No. 6.4e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAWSNDNIPYLDH 21
Db 481 ANDHLSILEAWSNDNIPYLDH 501

RESULT 2
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07

```

EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRF  
ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 100.0%; Score 119; DB 4; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 6,4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ANDHLSLEAWSNDPTPLHD 21  
|||||  
DB 481 ANDHLSLEAWSNDPTPLHD 501

RESULT 3  
US-09-210-361-4  
Sequence 4, Application US/09210361  
Patent No. 6284479  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starches and  
FILE REFERENCE: 0357CR  
CURRENT APPLICATION NUMBER: US/09/210,361  
EARLIER FILING DATE: 1998-12-11  
EARLIER APPLICATION NUMBER: 09/007,999  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/009,620  
EARLIER FILING DATE: 1998-01-20  
EARLIER APPLICATION NUMBER: 08/485,243  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 1375  
TYPE: PRF  
ORGANISM: streptococcus mutans  
US-09-210-361-4

Query Match 92.4%; Score 110; DB 4; Length 1375;  
Best Local Similarity 95.2%; Pred. No. 1.7e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANDHLSLEAWSNDPTPLHD 21  
|||||  
DB 507 ANDHLSLEAWSNDPTPLHD 527

RESULT 4  
US-09-008-172-2  
Sequence 2, Application US/09008172  
Patent No. 6127602  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starch and  
FILE REFERENCE: 0358D  
CURRENT APPLICATION NUMBER: US/09/008,172

CURRENT FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1430  
TYPE: PRF  
ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 63.0%; Score 75; DB 3; Length 1430;  
Best Local Similarity 71.4%; Pred. No. 0.00084;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSLEAWSNDPTPLHD 21  
|||||  
DB 495 AINHLISLEAWSNDPQYKND 515

RESULT 5  
US-09-210-361-6  
Sequence 6, Application US/09210361  
Patent No. 6284479  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starches and  
FILE REFERENCE: 0357CR  
CURRENT APPLICATION NUMBER: US/09/210,361  
EARLIER FILING DATE: 1998-12-11  
EARLIER APPLICATION NUMBER: 09/007,999  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/009,620  
EARLIER FILING DATE: 1998-01-20  
EARLIER APPLICATION NUMBER: 08/485,243  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 1430  
TYPE: PRF  
ORGANISM: streptococcus mutans  
US-09-210-361-6

Query Match 63.0%; Score 75; DB 4; Length 1430;  
Best Local Similarity 71.4%; Pred. No. 0.00084;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSLEAWSNDPTPLHD 21  
|||||  
DB 495 AINHLISLEAWSNDPQYKND 515

RESULT 6  
US-08-793-824-2  
Sequence 2, Application US/08793824  
Patent No. 5981838  
GENERAL INFORMATION:  
APPLICANT: Simpson, Christine Lynn  
APPLICANT: Giffard, Philip Morrison  
APPLICANT: Jacques, Nicholas Anthony  
TITLE OF INVENTION: Genetic Manipulation of Plants to  
INCREASE STORED CARBOHYDRATES  
NUMBER OF SEQUENCES: 2



; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Griffith Hack & Co  
; STREET: Level 8, 168 Walker Street  
; CITY: No. 5981838th Sydney  
; STATE: New South Wales  
; COUNTRY: Australia  
; ZIP: 2060  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/793,824  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: AU PM7643  
; FILING DATE: 24-AUG-1994  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 61 2 9957 5944  
; TELEFAX: 61 2 957 6288  
; TELEX: 26547  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1577 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus salivarius  
; US-08-793-824-2

Query Match 52.9%; Score 63; DB 2; Length 1577;  
Best Local Similarity 72.2%; Pred. No. 0.084;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 HLSILEMSNDPTPLHD 21  
|||||  
Db 591 HLSILEMSYNDHQYKND 608

RESULT 7  
US-08-361-920-23  
; Sequence 23, Application US/08361920  
; Patent No. 5457046  
; GENERAL INFORMATION:  
; APPLICANT: Woeldike, Helle F.  
; APPLICANT: Hagen, Frederick  
; APPLICANT: Hjort, Carsten M.  
; APPLICANT: Sven, Hastrup  
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 54570460 No. 5457046disk of No. 5457046th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/361,920  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION: 435

; APPLICATION NUMBER: US 07/940,860  
; FILING DATE: 28-OCT-1992  
; APPLICATION NUMBER: DK 1158/90  
; FILING DATE: 09-MAY-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK91/00124  
; FILING DATE: 08-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J.  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 3435.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-867-0298  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 385 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-361-920-23

Query Match 37.8%; Score 45; DB 1; Length 385;  
Best Local Similarity 24.4%; Pred. No. 13;  
Matches 10; Conservative 5; Mismatches 6; Indels 20; Gaps 1;

OY 1 ANDHLSLEA-----MSDNDPTPLHD 21  
|||:|:|:|  
Db 329 ANDYLTVMNACIAPKCVGITWVGVSDBKDSWRPGDNPPLYD 369

RESULT 8  
US-08-479-939-23  
; Sequence 23, Application US/08479939  
; Patent No. 5686593  
; GENERAL INFORMATION:  
; APPLICANT: Woeldike, Helle F.  
; APPLICANT: Hagen, Frederick  
; APPLICANT: Hjort, Carsten M.  
; APPLICANT: Sven, Hastrup  
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 56865930 No. 5686593disk of No. 5686593th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/479,939  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/361,920  
; FILING DATE: 22-DEC-1994  
; APPLICATION NUMBER: US 07/940,860  
; FILING DATE: 28-OCT-1992  
; APPLICATION NUMBER: DK 1158/90  
; FILING DATE: 09-MAY-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK91/00124  
; FILING DATE: 08-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J.  
; REGISTRATION NUMBER: 33,728

```

? REFERENCE/DOCKET NUMBER: 3435-204-US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212-867-0123
? TELEFAX: 212-867-0298
? INFORMATION FOR SEQ ID NO: 23:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 385 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? IS-08-479-939-23

```

Query Match	37.8%	Score 45;	DB 1;	Length 385;
Best Local Similarity	24.4%	Pred. No. 13;		
Matches 10;	Conservative 5;	Mismatches 6;	Indels 20;	Gaps 1

```

1  ANDHLSILEA-----WSDNDTPYLHD 21
   |||:::|         | | | |
329 ANDYLTVMNACLAVPKCVGITVWGVSDKDSMRPGDNPLLYD 365

```

```

1      RESULT 9
2      US-08-483-432-23
3      / Sequence 23, Application US/08483432
4      / Patent No. 5763254
5      / GENERAL INFORMATION:
6      / APPLICANT: Woeldike, Helle F.
7      / APPLICANT: Hagen, Frederick
8      / APPLICANT: Hjort, Carsten M.
9      / APPLICANT: Sven, Hastrup
10     / TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose
11     / TITLE OF INVENTION: or Hemicellulose
12     / NUMBER OF SEQUENCES: 85
13     / CORRESPONDENCE ADDRESS:
14     / ADDRESSEE: No. 57632540 No. 5763254disk of No. 5763254th America, Inc
15     / STREET: 405 Lexington Avenue, 62nd Floor
16     / CITY: New York
17     / STATE: New York
18     / COUNTRY: United States of America
19     / ZIP: 10174-6201
20     / COMPUTER READABLE FORM:
21     / MEDIUM TYPE: Floppy disk
22     / COMPUTER: IBM PC compatible
23     / OPERATING SYSTEM: PC-DOS/MS-DOS
24     / SOFTWARE: Patent in Release #1.0, Version #1.25
25     / CURRENT APPLICATION DATA:
26     / APPLICATION NUMBER: US/08/483,432
27     / FILING DATE: 07-JUN-1995
28     / CLASSIFICATION: 435
29     / PRIOR APPLICATION DATA:
30     / APPLICATION NUMBER: US/08/361,920
31     / FILING DATE:
32     / APPLICATION NUMBER: US/07/940,860
33     / FILING DATE: 28-OCT-1992
34     / APPLICATION NUMBER: DK 1158/90
35     / FILING DATE: 09-MAY-1990
36     / PRIOR APPLICATION DATA:
37     / APPLICATION NUMBER: PCT/DK91/00124
38     / FILING DATE: 08-MAY-1991
39     / ATTORNEY/AGENT INFORMATION:
40     / NAME: Lambiris, Elias J.
41     / REGISTRATION NUMBER: 33,728
42     / REFERENCE/DOCKET NUMBER: 3435.204-US
43     / TELECOMMUNICATION INFORMATION:
44     / TELEPHONE: 212-867-0123
45     / TELEFAX: 212-867-0298
46     / INFORMATION FOR SEQ ID NO: 23:
47     / SEQUENCE CHARACTERISTICS:
48     / LENGTH: 385 amino acids
49     / TYPE: amino acid
50     / TOPOLOGY: linear
51     / MOLECULE TYPE: protein

```

US-08-483-432-23

Query Match	37.8%;	Score 45;	DB 1;	Length 385;
Best Local Similarity	24.4%;	Pred. No. 13;		
Matches 10;	Conservative 5;	Mismatches 6;	Indels 20;	Gaps 1;

```

QY      1 ANDHSLLEA-----WSDNDTPYLHD 21
        |||:::|      | | | |
Db      329 ANDYLTFMNAELAVPKCVGITVGVGSDKDSWRPGDNPPLVD 362

```

```

RESULT 10
US-09-311-170-2
: Sequence 2, Application US/09311170
: Patent No. 6121034
: GENERAL INFORMATION:
: APPLICANT: Laroche et al., Andre L.
: TITLE OF INVENTION: xylanase cxy1
: FILE REFERENCE: xylanase cxy1
: CURRENT APPLICATION NUMBER: US/09/311,170
: CURRENT FILING DATE: 1999-05-13
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 384
: TYPE: PRF
: ORGANISM: Coniothyrium minitans
: US-09-311-170-2

```

Query Match	37.0%;	Score 44;	DB 3;	Length 384;
Best Local Similarity	22.0%;	Pred. No. 19;		
Matches	9;	Conservative	8;	Mismatches 4;
				Indels 20;
				Gaps 1;

```
QY      1 ANDHLSIL-----EAWSDNDTPYLHD 21
      |||:::
Db      328 ANDYLTVMNGCLAVPKCVGITVMGVSDDKSDSRSDSPLEED 368
```

RESULT 11  
 US-08-596-291-3  
 Sequence 3, Application US/08596291  
 Patent No. 5821075  
 GENERAL INFORMATION:  
 APPLICANT: GONEZ, LEONEL JORGE  
 APPLICANT: SARAS, JAN  
 APPLICANT: CLAESSEN-WELSH, LENA  
 APPLICANT: HELDIN, CARL-HENRIK  
 TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
 TITLE OF INVENTION: EXPRESSION OF NOCOTECTIDE SEQUENCES FOR NOVEL PROTEIN  
 NUMBER OF INVENTION: TYROSINE PHOSPHATASES  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
 STREET: 600 ATLANTIC AVENUE  
 CITY: BOSTON  
 STATE: MASSACHUSETTS  
 COUNTRY: USA  
 ZIP: 02210  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/596,291  
 FILING DATE: 09-AUG-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/115,573  
 FILING DATE: 01-SEP-1993  
 ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: L0461/7000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
TELEX: 92-1742 EZEKEL  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2465 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-596-291-3

Query Match 36.1%; Score 43; DB 2; Length 2465;  
Best Local Similarity 56.2%; Pred. No. 2.5e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

4 H1SILE--AWSNDTP 17  
1:1 11 11111  
2347 H1SHLNFTAMPDHP 2362

RESULT 12  
US-09-100-804-3  
Sequence 3, Application US/09100804  
Patent No. 6066472  
GENERAL INFORMATION:  
APPLICANT: GONZ, LEONEL JORGE  
APPLICANT: SARAS, JAN  
APPLICANT: CLAESON-WELSH, LENA  
APPLICANT: HELDIN, CARL-HENRIK  
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/100,804  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/596,291  
FILING DATE: 09-AUG-1996  
APPLICATION NUMBER: US 08/115,573  
FILING DATE: 01-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/09943  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: L0461/7003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2465 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-100-804-3

Query Match 36.1%; Score 43; DB 3; Length 2465;  
Best Local Similarity 56.2%; Pred. No. 2.5e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

4 H1SILE--AWSNDTP 17  
1:1 11 11111  
Db 2347 H1SHLNFTAMPDHP 2362

RESULT 13  
US-09-080-855-12  
Sequence 12, Application US/09080855A  
Patent No. 6083721  
GENERAL INFORMATION:  
APPLICANT: Saras, Jan  
APPLICANT: Franz, Petra  
APPLICANT: Aspenstrm, Pontus  
APPLICANT: Hellman, Ulf  
APPLICANT: Gonzalez, Leonel Jorge  
APPLICANT: Heldin, Carl-Henrik  
TITLE OF INVENTION: PARG, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTP1  
FILE REFERENCE: L0461/7030  
CURRENT APPLICATION NUMBER: US/09/080,855A  
CURRENT FILING DATE: 1998-05-18  
EARLIER APPLICATION NUMBER: 08/805,583  
EARLIER FILING DATE: 1997-02-25  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 2466  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-080-855-12

Query Match 36.1%; Score 43; DB 3; Length 2466;  
Best Local Similarity 56.2%; Pred. No. 2.5e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

4 H1SILE--AWSNDTP 17  
1:1 11 11111  
Db 2348 H1SHLNFTAMPDHP 2363

RESULT 14  
PCT-US94-09943-2  
Sequence 2, Application PC/TUS9409943  
GENERAL INFORMATION:  
APPLICANT:  
APPLICANT:  
APPLICANT:  
APPLICANT:  
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESSES:  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/09943
: FILING DATE: 01-SEP-1994
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/115,573
: FILING DATE: 01-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: TWOMEY, MICHAEL J.
: REGISTRATION NUMBER: P-38,349
: REFERENCE/DOCKET NUMBER: LC461/7000MO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/720-3500
: TELEFAX: 617/720-2441
: TELEX: 92-1742 EZEKIEL
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2466 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US94-09943-2

```

```

Query Match          36.1%; Score 43; DB 5; Length 2466;
Best Local Similarity 56.2%; Pred. No. 2.5e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

```

```

OY 4 HLSILE--AMSDNDTP 17
: | | | | | | | | | |
DB 2348 HISHLNTAMPDHDTP 2363

```

```

RESULT 15
US-09-290-640-46
: Sequence 46, Application US/09290640
: Patent No. 6204055
: GENERAL INFORMATION:
: APPLICANT: Dean, Nicholas M.
: TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling
: FILE REFERENCE: ISPH-0351
: CURRENT APPLICATION NUMBER: US/09/290,640
: CURRENT FILING DATE: 1999-04-12
: NUMBER OF SEQ ID NOS: 85
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 46
: LENGTH: 2485
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-290-640-46

```

```

Query Match          36.1%; Score 43; DB 4; Length 2485;
Best Local Similarity 56.2%; Pred. No. 2.5e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;
OY 4 HLSILE--AMSDNDTP 17
: | | | | | | | | | |
DB 2367 HISHLNTAMPDHDTP 2382

```

Search completed: March 27, 2002, 13:59:29  
Job time: 582 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:24 ; Search time 1139.61 Seconds

(without alignments)  
5.116 Million cell updates/sec

Title: US-09-290-049a-10

Sequence: 1 ANDHLSILEAMSDNDTPYLHD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

1 number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/PCRTUS.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US06.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US080.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US081.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US082.COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US083.COMB.pep.\*  
8: /cgn2\_6/ptodata/2/paa/US084.COMB.pep.\*  
9: /cgn2\_6/ptodata/2/paa/US085.COMB.pep.\*  
10: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
11: /cgn2\_6/ptodata/2/paa/US087.COMB.pep.\*  
12: /cgn2\_6/ptodata/2/paa/US088.COMB.pep.\*  
13: /cgn2\_6/ptodata/2/paa/US089.COMB.pep.\*  
14: /cgn2\_6/ptodata/2/paa/US090.COMB.pep.\*  
15: /cgn2\_6/ptodata/2/paa/US091.COMB.pep.\*  
16: /cgn2\_6/ptodata/2/paa/US092.COMB.pep.\*  
17: /cgn2\_6/ptodata/2/paa/US093.COMB.pep.\*  
18: /cgn2\_6/ptodata/2/paa/US094.COMB.pep.\*  
19: /cgn2\_6/ptodata/2/paa/US095.COMB.pep.\*  
20: /cgn2\_6/ptodata/2/paa/US096.COMB.pep.\*  
21: /cgn2\_6/ptodata/2/paa/US097.COMB.pep.\*  
22: /cgn2\_6/ptodata/2/paa/US098.COMB.pep.\*  
23: /cgn2\_6/ptodata/2/paa/US099.COMB.pep.\*  
24: /cgn2\_6/ptodata/2/paa/US60.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	21	US-09-290-049-1	Sequence 1, Appl1
2	119	100.0	21	US-09-290-049-10	Sequence 10, Appl1
3	119	100.0	1475	US-09-557-848-2	Sequence 2, Appl1
4	119	100.0	1475	US-09-740-274-2	Sequence 2, Appl1
5	110	92.4	1375	US-09-740-274-4	Sequence 4, Appl1
6	108	90.8	21	US-09-290-049-12	Sequence 12, Appl1
7	108	90.8	21	US-09-290-049-14	Sequence 14, Appl1
8	75	63.0	21	US-09-290-049-11	Sequence 11, Appl1
9	75	63.0	1430	US-09-649-885-2	Sequence 2, Appl1

10	75	63.0	1430	21	US-09-740-274-6	Sequence 6, Appl1
11	73	61.3	21	16	US-09-290-049-13	Sequence 13, Appl1
12	60	50.4	2057	18	US-09-499-203-2	Sequence 2, Appl1
13	50	42.0	306	19	US-09-595-298A-661	Sequence 661, App
14	50	42.0	340	19	US-09-595-298A-660	Sequence 660, App
15	49.5	41.6	546	24	US-60-324-109-17015	Sequence 17015, A
16	49.5	41.6	549	24	US-60-324-109-16789	Sequence 16789, A
17	49	41.2	391	24	US-60-324-109-20736	Sequence 20736, A
18	48.5	40.8	401	1	PCT-US01-01309-222	Sequence 222, App
19	48	40.3	773	16	US-09-248-796-22573	Sequence 22573, A
20	47.5	39.9	178	22	US-09-867-716-19848	Sequence 19848, A
21	47.5	39.9	534	24	US-09-344-882-24	Sequence 24, Appl1
22	47.5	39.9	534	24	US-60-324-109-33149	Sequence 33149, A
23	47	39.5	196	19	US-09-595-298A-765	Sequence 765, App
24	46.5	39.1	220	24	US-60-212-413-164	Sequence 164, App
25	46.5	39.1	220	24	US-60-229-518-331	Sequence 331, App
26	46.5	39.1	330	24	US-60-207-422-75	Sequence 75, Appl1
27	46.5	39.1	539	1	PCT-US01-01332-762	Sequence 762, App
28	46.5	39.1	539	21	US-09-764-875-762	Sequence 762, App
29	46	38.7	62	18	US-09-417-507-26597	Sequence 26597, A
30	46	38.7	94	1	PCT-US01-08656-5595	Sequence 5595, A
31	46	38.7	118	20	US-09-673-395A-221	Sequence 221, App
32	46	38.7	199	21	US-09-738-626-4571	Sequence 4571, App
33	46	38.7	275	19	US-09-538-092-348	Sequence 348, App
34	46	38.7	336	15	US-09-198-452A-662	Sequence 662, App
35	46	38.7	336	18	US-09-438-185-626	Sequence 626, App
36	46	38.7	394	21	US-09-739-449-8822	Sequence 8822, App
37	46	38.7	394	22	US-09-803-110-8822	Sequence 8822, App
38	46	38.7	476	1	PCT-US01-122636-9	Sequence 9, Appl1
39	46	38.7	476	24	US-60-219-231-9	Sequence 9, Appl1
40	46	38.7	596	16	US-09-252-991A-21255	Sequence 21255, A
41	46	38.7	2590	24	US-60-215-161-7203	Sequence 7203, App
42	46	38.7	2603	24	US-60-215-161-5891	Sequence 5891, App
43	45	37.8	185	24	US-60-215-161-5719	Sequence 5719, App
44	45	37.8	194	1	PCT-US01-14827-9825	Sequence 9825, App
45	45	37.8	236	15	US-09-154-750-80	Sequence 80, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-290-049-1  
Sequence 1, Application US/09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taudman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: PDC98-0102A  
CURRENT APPLICATION NUMBER: US/09/290,049  
CURRENT FILING DATE: 1999-04-12  
EARLIER APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115,142  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURES:  
OTHER INFORMATION: EAW peptide  
US-09-290-049-1

Query Match 100.0%; Score 119; DB 16; Length 21;  
Best Local Similarity 100.0%; Pred. No. 5.3e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ANDHLSILEAMSDNDTPYLHD 21  
|||||

Db 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 2

US-09-290-049-10

Sequence 10, Application US/09290049

GENERAL INFORMATION:

APPLICANT: Smith, Daniel J.

APPLICANT: Traubman, Martin A.

TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

FILE REFERENCE: FC98-01P2A

CURRENT APPLICATION NUMBER: US/09/290,049

CURRENT FILING DATE: 1999-04-12

EARLIER APPLICATION NUMBER: 60/081,550

EARLIER FILING DATE: 1998-04-13

EARLIER APPLICATION NUMBER: 60/115,142

EARLIER FILING DATE: 1999-01-08

NUMBER OF SEQ ID NOS: 19

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 10

LENGTH: 21

TYPE: PRT

ORGANISM: S. mutans

US-09-290-049-10

Query Match 100.0%; Score 119; DB 16; Length 21;

Best Local Similarity 100.0%; Pred. No. 5.3e-11;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDTPYLHD 21

Db 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 3

US-09-557-848-2

Sequence 2, Application US/09557848

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Substitutes for Modified Starch and

FILE REFERENCE: 0356D2

CURRENT APPLICATION NUMBER: US/09/557,848

CURRENT FILING DATE: 2000-04-26

EARLIER APPLICATION NUMBER: 08/478,704

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 09/007,999

EARLIER FILING DATE: 1998-01-16

NUMBER OF SEQ ID NOS: 2

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 1475

TYPE: PRT

ORGANISM: Streptococcus mutans

US-09-557-848-2

Query Match 100.0%; Score 119; DB 19; Length 1475;

Best Local Similarity 100.0%; Pred. No. 5.2e-09;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDTPYLHD 21

Db 481 ANDHLSILEAMSDNDTPYLHD 501

RESULT 4

US-09-740-274-2

Sequence 2, Application US/09740274

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

Query Match 100.0%; Score 119; DB 21; Length 1475;

Best Local Similarity 100.0%; Pred. No. 5.2e-09;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDTPYLHD 21

Db 481 ANDHLSILEAMSDNDTPYLHD 501

RESULT 5

US-09-740-274-4

Sequence 4, Application US/09740274

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Glucan-containing Compositions and Paper

FILE REFERENCE: 0357CRD

CURRENT APPLICATION NUMBER: US/09/740,274

CURRENT FILING DATE: 2000-12-19

PRIOR APPLICATION NUMBER: 09/210,361

PRIOR FILING DATE: 1998-12-11

PRIOR APPLICATION NUMBER: 09/007,999

PRIOR FILING DATE: 1998-01-16

PRIOR APPLICATION NUMBER: 08/478,704

PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: 09/009,620

PRIOR FILING DATE: 1998-01-20

PRIOR APPLICATION NUMBER: 08/485,243

PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: 09/008,172

PRIOR FILING DATE: 1998-01-16

PRIOR APPLICATION NUMBER: 08/482,711

PRIOR FILING DATE: 1995-06-07

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4

LENGTH: 1375

TYPE: PRT

ORGANISM: Streptococcus mutans

US-09-740-274-4

Query Match 92.4%; Score 110; DB 21; Length 1375;

Best Local Similarity 95.2%; Pred. No. 1.2e-07;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDTPYLHD 21

Db 507 ANDHLSILEAMSDNDPTYLHD 527

RESULT 6

US-09-290-049-12

; Sequence 12, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT FILING DATE: 1999-04-12

; EARLIER FILING DATE: 1998-04-13

; EARLIER FILING DATE: 1998-04-13

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 12

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. downei

US-09-290-049-12

Query Match 90.8%; Score 108; DB 16; Length 21;  
Best Local Similarity 85.7%; Pred. No. 2.7e-09;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANDHLSILEAMSDNDPTYLHD 21

Db 1 ANNHVSIVEAMSDNDPTYLHD 21

RESULT 7

US-09-290-049-14

; Sequence 14, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT FILING DATE: 1999-04-12

; EARLIER FILING DATE: 1998-04-13

; EARLIER FILING DATE: 1998-04-13

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 14

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. sobrinus

US-09-290-049-14

Query Match 90.8%; Score 108; DB 16; Length 21;  
Best Local Similarity 85.7%; Pred. No. 2.7e-09;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANDHLSILEAMSDNDPTYLHD 21

Db 1 ANNHVSIVEAMSDNDPTYLHD 21

RESULT 8

US-09-290-049-11

; Sequence 11, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT FILING DATE: 1999-04-12

; EARLIER FILING DATE: 1998-04-13

; EARLIER FILING DATE: 1998-04-13

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 11

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. mutans

US-09-290-049-11

Query Match 63.0%; Score 75; DB 16; Length 21;  
Best Local Similarity 71.4%; Pred. No. 0.00036;

Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ANDHLSILEAMSDNDPTYLHD 21

Db 1 AINHLISLEAMSDNDPQYKND 21

RESULT 9

US-09-649-885-2

; Sequence 2, Application US/09649885

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Substitutes for Modified Starch and

; FILE REFERENCE: 0358D2

; CURRENT FILING DATE: 2000-08-28

; PRIOR FILING DATE: 1998-01-16

; PRIOR FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 1430

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-649-885-2

Query Match 63.0%; Score 75; DB 20; Length 1430;  
Best Local Similarity 71.4%; Pred. No. 0.034;

Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ANDHLSILEAMSDNDPTYLHD 21

Db 495 AINHLISLEAMSDNDPQYKND 515

RESULT 10

US-09-740-274-6

; Sequence 6, Application US/09740274

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Glucan-containing Compositions and Paper

; FILE REFERENCE: 0357CRD

; CURRENT FILING DATE: 2000-12-19

; PRIOR FILING DATE: 1998-12-11

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 1430

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-740-274-6

PRIOR APPLICATION NUMBER: 09/007,999  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: streptococcus mutans  
09-740-274-6

Query Match 63.0%; Score 75; DB 21; Length 1430;  
Best Local Similarity 71.4%; Pred. No. 0.034;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSNDPTPLHD 21  
|:|||||||||  
DB 495 AINHLSLEAMSNDPDYKND 515

RESULT 11  
US-09-290-049-13  
; Sequence 13, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01p2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 13  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049-13

Query Match 61.3%; Score 73; DB 16; Length 21;  
Best Local Similarity 73.7%; Pred. No. 0.00073;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSNDPTPL 19  
|:|||||||||  
DB 1 AIDHLSLEAMSGNDNDYV 19

RESULT 12  
US-09-499-203-2  
; Sequence 2, Application US/09499203  
; GENERAL INFORMATION:  
; APPLICANT: KOSSMANN, Jens  
; APPLICANT: WELSH, Thomas  
; APPLICANT: QUANZ, Martin  
; APPLICANT: KNOTH, Karola  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
; FILE REFERENCE: 147-196p

CURRENT APPLICATION NUMBER: US/09/499,203  
CURRENT FILING DATE: 2000-02-08  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 2057  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-499-203-2

Query Match 50.4%; Score 60; DB 18; Length 2057;  
Best Local Similarity 55.0%; Pred. No. 11;  
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSNDPTPLH 20  
|:|||||  
DB 665 ANKHLSLEDMNGKDPQYVN 684

RESULT 13  
US-09-595-298a-661  
; Sequence 661, Application US/09595298a  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai  
; APPLICANT: BROVER, Vyacheslav  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept  
; FILE REFERENCE: 2750-0953p  
; CURRENT APPLICATION NUMBER: US/09/595,298a  
; CURRENT FILING DATE: 2000-06-16  
; NUMBER OF SEQ ID NOS: 2756  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 661  
; LENGTH: 306  
; TYPE: PRT  
; ORGANISM: Arabidopsis Thaliana  
; FEATURE:  
; NAME/KEY: peptide  
; LOCATION: (1)..(306)  
; OTHER INFORMATION: Ceres Seq. ID no. 1026741  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: Xaa is any aa, unknown or other  
US-09-595-298a-661

Query Match 42.0%; Score 50; DB 19; Length 306;  
Best Local Similarity 52.9%; Pred. No. 49;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DHLSLEAMSNDPTPL 19  
|:|||||  
DB 200 DYEOLMEAMSDKGLTV 216

RESULT 14  
US-09-595-298a-660  
; Sequence 660, Application US/09595298a  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai  
; APPLICANT: BROVER, Vyacheslav  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept  
; FILE REFERENCE: 2750-0953p  
; CURRENT APPLICATION NUMBER: US/09/595,298a  
; CURRENT FILING DATE: 2000-06-16  
; NUMBER OF SEQ ID NOS: 2756  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 660  
; LENGTH: 340  
; TYPE: PRT  
; ORGANISM: Arabidopsis Thaliana



```

FEATURE:
; NAME/KEY: peptide
; LOCATION: (1)..(340)
; OTHER INFORMATION: Ceres Seq. ID no. 1026740
; NAME/KEY: misc_feature
; LOCATION: ()..()
; OTHER INFORMATION: Xaa is any aa, unknown or other
US-09-595-298A-660

```

```

Query Match          42.0%; Score 50; DB 19; Length 340;
Best Local Similarity 52.9%; Pred. No. 55;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      3 DHLTLEAWSNDPTPYL 19
         | : | | | | | | | :
Db      234 DYEQIMEAWSDKGTLYV 250

```

# RESULT 15

```

US-60-324-109-17015
Sequence 17015, Application US/60324109
GENERAL INFORMATION:

```

```

; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jindong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)B
; CURRENT APPLICATION NUMBER: US/60/324,109
; CURRENT FILING DATE: 2001-09-21
; NUMBER OF SEQ ID NOS: 33196
; SEQ ID NO 17015
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
US-60-324-109-17015

```

```

Query Match          41.6%; Score 49.5; DB 24; Length 546;
Best Local Similarity 58.8%; Pred. No. 1.1e+02;
Matches 10; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

```

```

QY      2 NDHLSILEAWSNDPTPY 18
         | | : | | | | | | |
Db      144 NDELALLETW-DNGKPY 159

```

```

Search completed: March 27, 2002, 14:20:24
Job time: 1571 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:46 ; Search time 137.48 seconds  
(without alignments)  
10.540 Million cell updates/sec

Title: US-09-290-049a-10  
Perfect score: 119  
Sequence: 1 ANDHLSILEAWSNDNPTLYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 6899538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCr\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	21	US-09-562-328-20	Sequence 20, Appl
2	119	100.0	21	US-09-290-049a-1	Sequence 1, Appl
3	119	100.0	21	US-09-290-049a-10	Sequence 10, Appl
4	110	92.4	21	US-09-562-328-21	Sequence 21, Appl
5	108	90.8	21	US-09-562-328-23	Sequence 23, Appl
6	108	90.8	21	US-09-562-328-25	Sequence 25, Appl
7	108	90.8	21	US-09-290-049a-12	Sequence 12, Appl
8	108	90.8	21	US-09-290-049a-14	Sequence 14, Appl
9	75	63.0	21	US-09-562-328-22	Sequence 22, Appl
10	75	63.0	21	US-09-290-049a-11	Sequence 11, Appl
11	75	63.0	21	US-09-604-957-4	Sequence 4, Appl
12	73	61.3	21	US-09-562-328-24	Sequence 24, Appl
13	73	61.3	21	US-09-290-049a-13	Sequence 13, Appl
14	71	59.7	21	US-09-604-957-5	Sequence 5, Appl
15	60	50.4	21	US-09-604-957-6	Sequence 6, Appl
16	49.5	41.6	423	US-09-708-427-55326	Sequence 55326, A
17	49.5	41.6	511	US-09-708-427-55325	Sequence 55325, A
18	49.5	41.6	549	US-09-708-427-55324	Sequence 55324, A
19	49	41.2	302	PCT-US02-03987-14004	Sequence 14004, A
20	49	41.2	302	US-09-815-242-14004	Sequence 14004, A
21	49	41.2	302	US-10-072-851-14004	Sequence 14004, A
22	49	41.2	535	US-09-604-957-7	Sequence 7, Appl
23	49	41.2	1278	US-09-604-957-3	Sequence 3, Appl
24	47.5	39.9	434	US-09-708-427-7994	Sequence 7994, Ap
25	47.5	39.9	519	US-09-708-427-7993	Sequence 7993, Ap

26	47	39.5	15	7	US-10-067-649-90	Sequence 90, Appl
27	47	39.5	366	6	US-09-708-427-27415	Sequence 27415, A
28	47	39.5	367	6	US-09-708-427-27416	Sequence 27416, A
29	47	39.5	378	6	US-09-708-427-27414	Sequence 27414, A
30	46.5	39.1	1604	6	US-09-888-615-73	Sequence 73, Appl
31	46	38.7	199	6	US-09-605-703B-2226	Sequence 2226, Ap
32	46	38.7	2590	6	US-09-897-516-5891	Sequence 5891, Ap
33	46	38.7	2603	6	US-09-897-516-5891	Sequence 5891, Ap
34	45	37.8	185	6	US-09-897-516-5719	Sequence 5719, Ap
35	45	37.8	236	6	US-09-154-750A-80	Sequence 80, Appl
36	45	37.8	638	6	US-09-708-427-25322	Sequence 25322, A
37	45	37.8	765	6	US-09-708-427-25321	Sequence 25321, A
38	45	37.8	947	6	US-09-708-427-25320	Sequence 25320, A
39	44.5	37.4	499	6	PCT-US02-03987-11634	Sequence 42525, A
40	44	37.0	322	1	US-09-815-242-11634	Sequence 11634, A
41	44	37.0	322	7	US-10-072-851-11634	Sequence 11634, A
42	44	37.0	331	1	PCT-US02-03987-10188	Sequence 10188, A
43	44	37.0	331	6	US-09-815-242-10188	Sequence 10188, A
44	44	37.0	331	7	US-10-072-851-10188	Sequence 10188, A
45	44	37.0	331	7	US-10-072-851-10188	Sequence 10188, A

#### ALIGNMENTS

RESULT 1  
US-09-562-328-20  
; Sequence 20, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEE, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562.328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288.965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-20

Query Match 100.0%; Score 119; DB 6; Length 21;  
Best Local Sim. Early 100.0%; Pred. No. 2.9e-12;  
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ANDHLSILEAWSNDNPTLYLHD 21  
DB 1 ANDHLSILEAWSNDNPTLYLHD 21  
RESULT 2  
US-09-290-049a-1  
; Sequence 1, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290.049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081.550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115.142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0

```
SEQ ID NO 1
LENGTH: 21
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: EAW peptide
US-09-290-049A-1
```

```
Query Match      100.0%; Score 119; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSLEAMSNDPTPYLHD 21
      |||
Db      1 ANDHLSLEAMSNDPTPYLHD 21
```

```
US-09-290-049A-10
```

```
Sequence 10, Application US/09290049A
GENERAL INFORMATION:
APPLICANT: Smith, Daniel J.
APPLICANT: Taubman, Martin A.
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
FILE REFERENCE: 1564,1008-002
CURRENT APPLICATION NUMBER: US/09/290,049A
CURRENT FILING DATE: 1999-04-12
PRIOR APPLICATION NUMBER: 60/081,550
PRIOR FILING DATE: 1998-04-13
PRIOR APPLICATION NUMBER: 60/115,142
PRIOR FILING DATE: 1999-01-08
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10
LENGTH: 21
TYPE: PRT
ORGANISM: S. mutans
US-09-290-049A-10
```

```
Query Match      100.0%; Score 119; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
1 ANDHLSLEAMSNDPTPYLHD 21
|||
Db      1 ANDHLSLEAMSNDPTPYLHD 21
```

```
US-09-562-328-21
```

```
Sequence 21, Application US/09562328
GENERAL INFORMATION:
APPLICANT: LEES, ANDREW
APPLICANT: TAUBMAN, MARTIN A.
APPLICANT: SMITH, DANIEL J.
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
FILE REFERENCE: 04995,0046-01
CURRENT APPLICATION NUMBER: US/09/562,328
CURRENT FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 09/288,965
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 21
LENGTH: 21
TYPE: PRT
ORGANISM: Streptococcus sp.
US-09-562-328-21
```

```
Query Match      92.4%; Score 110; DB 6; Length 21;
Best Local Similarity 95.2%; Pred. No. 7.4e-11;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSLEAMSNDPTPYLHD 21
      |||
Db      1 ANDHLSLEAMSNDPTPYLHD 21
```

```
US-09-562-328-23
```

```
Sequence 23, Application US/09562328
GENERAL INFORMATION:
APPLICANT: LEES, ANDREW
APPLICANT: TAUBMAN, MARTIN A.
APPLICANT: SMITH, DANIEL J.
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
FILE REFERENCE: 04995,0046-01
CURRENT APPLICATION NUMBER: US/09/562,328
CURRENT FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 09/288,965
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 23
LENGTH: 21
TYPE: PRT
ORGANISM: Streptococcus sp.
US-09-562-328-23
```

```
Query Match      90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSLEAMSNDPTPYLHD 21
      |||
Db      1 ANNHVSIVEAMSNDPTPYLHD 21
```

```
US-09-562-328-25
```

```
Sequence 25, Application US/09562328
GENERAL INFORMATION:
APPLICANT: LEES, ANDREW
APPLICANT: TAUBMAN, MARTIN A.
APPLICANT: SMITH, DANIEL J.
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
FILE REFERENCE: 04995,0046-01
CURRENT APPLICATION NUMBER: US/09/562,328
CURRENT FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 09/288,965
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 25
LENGTH: 21
TYPE: PRT
ORGANISM: Streptococcus sp.
US-09-562-328-25
```

```
Query Match      90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSLEAMSNDPTPYLHD 21
      |||
Db      1 ANNHVSIVEAMSNDPTPYLHD 21
```

```
US-09-562-328-12
```

```
; Sequence 12, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 21
; TYPE: PRF
; ORGANISM: S. downei
US-09-290-049A-12

Query Match          90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHSLLEAWSNDNDPYLHD 21
Db 1 AINHSLLEAWSNDNDPYLHD 21

RESULT 8
US-09-290-049A-14
; Sequence 14, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 21
; TYPE: PRF
; ORGANISM: S. sobrinus
US-09-290-049A-14

Query Match          90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHSLLEAWSNDNDPYLHD 21
Db 1 AINHSLLEAWSNDNDPYLHD 21

RESULT 9
US-09-562-328-22
; Sequence 22, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
```

```
; CURRENT APPLICATION NUMBER: US/09/562.328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 21
; TYPE: PRF
; ORGANISM: Streptococcus sp.
US-09-562-328-22

Query Match          63.0%; Score 75; DB 6; Length 21;
Best Local Similarity 71.4%; Pred. No. 2.3e-05;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHSLLEAWSNDNDPYLHD 21
Db 1 AINHSLLEAWSNDNDPYLHD 21

RESULT 10
US-09-290-049A-11
; Sequence 11, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRF
; ORGANISM: S. mutans
US-09-290-049A-11

Query Match          63.0%; Score 75; DB 6; Length 21;
Best Local Similarity 71.4%; Pred. No. 2.3e-05;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHSLLEAWSNDNDPYLHD 21
Db 1 AINHSLLEAWSNDNDPYLHD 21

RESULT 11
US-09-604-957-6
; Sequence 4, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUTI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604.957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
```

Search completed: March 27, 2002, 14:22:47  
Job time: 1694 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:17 ; Search time 102.51 Seconds

(without alignments)  
15.605 Million cell updates/sec

Title: US-09-290-049a-10

Perfect score: 119

Sequence: 1 ANDHLSILEAWSNDPTPLHD 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	1475	2 B33135	gtfB protein precu
2	110	92.4	1375	2 UT0345	dextranucrase (EC
3	108	90.8	1592	2 A38175	glucosyltransferas
4	78	65.0	1365	2 A41483	glucosyltransferas
5	75	63.0	1431	2 A45866	dextranucrase (EC
6	74	62.2	1508	2 T31098	probable dextranu
7	66	55.5	1449	2 T30857	glucosyltransferas
8	66	55.5	1449	2 T30552	glucosyltransferas
9	63	52.9	1577	2 T30858	glucosyltransferas
10	58	48.7	1518	2 A44811	glucosyltransferas
11	57	47.9	331	2 B48445	glyceraldhyde-3-p
12	57	47.9	1599	2 S22737	glucosyltransferas
13	52	43.7	4848	2 T30289	pistinnamycin I sy
14	50	42.0	347	2 T48610	hypothetical prote
15	49.5	41.6	549	2 T03963	rt2 nuclear restor
16	48	40.3	336	1 H05E40	BGLF2 protein - hu
17	48	40.3	490	2 H05E40	probable ppdk prot
18	47.5	39.9	331	1 DEUTC	glyceraldhyde-3-p
19	47.5	39.9	519	2 C86372	hypothetical prote
20	47	39.5	175	2 C86205	hypothetical prote
21	47	39.5	378	2 T04254	hypothetical prote
22	47	39.5	524	2 D82220	conserved hypotet
23	47	39.5	525	2 T40088	RhoGEF domain cont
24	46	38.7	275	2 S55978	hypothetical prote
25	46	38.7	335	2 B86568	glyceraldhyde-3-p
26	46	38.7	335	2 B72053	glyceraldhyde-3-p
27	46	38.7	335	2 S58229	saliolate biosynt
28	45.5	38.2	96	2 E81786	conserved hypotet
29	45	37.8	236	2 S62732	guanidinacetate N

30	45	37.8	947	2 E86362	hypothetical prote
31	45	37.8	1159	2 T43461	probable phosphodi
32	44.5	37.4	210	2 H83332	conserved hypotet
33	44	37.0	247	2 PQ0178	glyceraldhyde-3-p
34	44	37.0	331	2 DEECG3	glyceraldhyde-3-p
35	44	37.0	331	2 D85788	glyceraldhyde-3-p
36	44	37.0	333	2 DECHG3	glyceraldhyde-3-p
37	44	37.0	333	2 JC5370	glyceraldhyde-3-p
38	44	37.0	337	1 DEJUC	glyceraldhyde-3-p
39	44	37.0	337	1 DEJUC	glyceraldhyde-3-p
40	44	37.0	337	1 DEJUC	glyceraldhyde-3-p
41	44	37.0	436	2 T24953	hypothetical prote
42	44	37.0	468	2 S61964	probable membrane
43	44	37.0	1208	2 B82091	exodeoxyribonuclea
44	44	37.0	1313	2 T29027	hypothetical prote
45	43.5	36.6	97	2 S12958	tachyinnin delta p

#### ALIGNMENTS

```

RESULT 1
B33135
gtfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 15-Oct-1999
C:Accession: B33135; A33128
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013
A:Accession: B33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SH1>
A:Cross-references: GB:M17361; NID:G153639; PIDN:AAA8588.1; PID:G153640
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
submitted to the Protein Sequence Database, September 1990
A:Reference number: A33128
A:Accession: A33128
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-171,173-641,'N',643-1475 <SH2>
A:Experimental source: strain GS-5
A:Superfamily: cpl repeat homology
F:1096-1115/Domain: cpl repeat homology <CP1>
F:1224-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>
F:1354-1373/Domain: cpl repeat homology <CP4>
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 100.0%; Score 119; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 3.9e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDPTPLHD 21
DB 481 ANDHLSILEAWSNDPTPLHD 501

RESULT 2
UT0345
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999
C:Accession: UT0345; C33135
R:Ueda, S.; Shiroza, T.; Kuramitsu, H. K.
Gene 69, 101-109, 1988
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
A:Reference number: UT0345; MUID:89137980
A:Accession: UT0345

```

A.Molecule type: DNA  
A.Residues: 1-1375 <UED>  
A.Experimental source: GS-5  
R.Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A.Title: Sequence analysis of the gltF gene from Streptococcus mutans.  
A.Reference number: A3135; MUID:87308013  
A.Accession: C3135  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-349 <SHI>  
A.Cross-references: GB:M17361  
C.Genetics:  
A.Gene: gltF  
C.Function:  
A.Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
C.Superfamily: cpl repeat homology  
Keywords: duplication; glycosyltransferase; hexosyltransferase  
-34/Domain: signal sequence #status predicted <SIG>  
-35-1375/Product: glucosyltransferase #status predicted <MAT>  
F.1126-1145/Domain: cpl repeat homology <CP1>  
F.1253-1272/Domain: cpl repeat homology <CP2>  
F.1318-1337/Domain: cpl repeat homology <CP3>

Query Match 92.4%; Score 110; DB 2; Length 1375;  
Best Local Similarity 95.2%; Pred. No. 9, 1e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHSLILEAWSNDTPYLHD 21  
DB 507 ANDHSLILEAWSNDTPYLHD 527

RESULT 3  
A38175  
glucosyltransferase precursor - Streptococcus sobrinus  
C.Species: Streptococcus sobrinus  
C.Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
C.Accession: A38175  
R.Abo, H.; Mtsunura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A.Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus  
A.Reference number: A38175; MUID:91123227  
A.Accession: A38175  
A.Status: preliminary  
Molecule type: DNA  
Residues: 1-1592 <ABO>  
A.Cross-references: GB:D90213; NID:9217032; PIDN:BAAL4241.1; PID:d1014946; PID:9217033  
C.Superfamily: cpl repeat homology  
F.1093-1112/Domain: cpl repeat homology <CP1>  
F.1222-1241/Domain: cpl repeat homology <CP2>  
F.1287-1306/Domain: cpl repeat homology <CP3>  
F.1330-1351/Domain: cpl repeat homology <CP4>  
F.1352-1371/Domain: cpl repeat homology <CP5>  
F.1402-1420/Domain: cpl repeat homology <CP6>  
F.1465-1484/Domain: cpl repeat homology <CP7>  
F.1513-1532/Domain: cpl repeat homology <CP8>

Query Match 90.8%; Score 108; DB 2; Length 1592;  
Best Local Similarity 85.7%; Pred. No. 2, 2e-08;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHSLILEAWSNDTPYLHD 21  
DB 477 ANNHVSIVEAWSNDTPYLHD 497

RESULT 4  
A41483  
glucosyltransferase (EC 2.4.1.-) gltS precursor - Streptococcus sobrinus  
C.Species: Streptococcus sobrinus

C.Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C.Accession: A41483  
R.Gilmore, K.S.; Russell, R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A.Title: Analysis of the Streptococcus downei gltS gene, which specifies a glucosyltr  
A.Reference number: A41483; MUID:90316665  
A.Accession: A41483  
A.Molecule type: DNA  
A.Residues: 1-1365 <GIL>  
A.Cross-references: GB:M30943; NID:9153652; PIDN:AAA26898.1; PID:9153653  
C.Genetics:  
A.Gene: gltS  
C.Superfamily: cpl repeat homology  
C.Keywords: glycosyltransferase; hexosyltransferase

Query Match 65.5%; Score 78; DB 2; Length 1365;  
Best Local Similarity 71.4%; Pred. No. 0.00087;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHSLILEAWSNDTPYLHD 21  
DB 467 AIDHSLILEAWSNDTPYLHD 487

RESULT 5  
A45866  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans  
C.Species: Streptococcus mutans  
C.Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C.Accession: A45866  
R.Honda, O.; Kato, C.; Kuramitsu, H.K.  
J. Gen. Microbiol. 136, 2099-2105, 1990  
A.Title: Nucleotide sequence of the Streptococcus mutans gltD gene encoding the gluco  
A.Reference number: A45866; MUID:91100958  
A.Accession: A45866  
A.Status: preliminary  
A.Molecule type: DNA  
Residues: 1-1431 <HON>  
A.Cross-references: GB:M29296  
C.Superfamily: cpl repeat homology  
C.Keywords: glycosyltransferase; hexosyltransferase  
F.181-201/Domain: cpl repeat homology <CP1>  
F.1127-1146/Domain: cpl repeat homology <CP2>  
F.1192-1211/Domain: cpl repeat homology <CP3>  
F.1257-1276/Domain: cpl repeat homology <CP4>  
F.1327-1397/Domain: cpl repeat homology <CP5>  
F.1321-1340/Domain: cpl repeat homology <CP6>  
F.1341-1361/Domain: cpl repeat homology <CP7>  
F.1385-1404/Domain: cpl repeat homology <CP8>

Query Match 63.0%; Score 75; DB 2; Length 1431;  
Best Local Similarity 71.4%; Pred. No. 0.0027;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHSLILEAWSNDTPYLHD 21  
DB 495 AINHLILEAWSNDTPYLHD 515

RESULT 6  
T31098  
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides  
C.Species: Leuconostoc mesenteroides  
C.Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C.Accession: T31098  
R.Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willems, R.M.  
FEMS Microbiol. Lett. 159, 307-315, 1998  
A.Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase  
A.Reference number: Z20981; MUID:98164374  
A.Accession: T31098  
A.Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-1508 <MON>  
A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AA895453.1  
A:Experimental source: strain NRRL B-1239  
C:Genetics:  
A:Gene: dsrB  
C:Function:  
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match	62.28;	Score 74;	DB 2;	Length 1508;
Best Local Similarity	66.78;	Pred. No. 0.0041;		
Matches 14; Conservative	1;	Mismatches 6;	Indels 0;	Gaps 0;

```
QY      1 ANDHLSILEAWSNDNDTPYIHD 21
        || ||||| | | | : |
Db      563 ANQHLSILEAWSNDNDPEYVKD 583
```

RESULT 7  
857

```

C:cosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: J30857
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pr
A:Reference number: J20909; MUID:95122197
A:Accession: J30857
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
C:Cross-references: EMBL.LJ35495; NID:J662378; PID:J662379; PIDN:AA41412.1
C:Genetics:
:Gene: gtlL

```

Query Match	55.5%	Score 66:	DB 2:	Length 1449:
Best Local Similarity	72.2%	Pred. No.	0.069	
Matches 13, Conservative	0:	Mismatches	5:	Indels 0:
				Gaps 0:

```

OY      4  HLSLEAWSNDNTPYLHD  21
          |||||
Db      539 HLSLEAWSHNDAYNED  556

```

RESULT 8  
30552

A:Accession: F30552  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1449 <JAF>  
 A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1  
 C:Genetics:  
 C:Gene: gtfN  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C:Accession: F30552  
 R:Jaife, R.I.  
 Submitted to the EMBL Data Library, February 1998  
 A:Description: Streptococcus salivarius V1477 gtfN.  
 A:Reference number: 220854

```

Query Match          55.5%;   Score 66;   DB 2;   Length 1449;
Best Local Similarity 72.2%;   Pred. No. 0.069;
Matches 13;   Conservative 0;   Mismatches 5;   Indels 0;   Gaps 0;

QY      4 HUSTLEAMSNDNTPYLHD 21
||||||| 11 | 1

```

Db 539 HLSILEAWSHNDAYNED 556

RESULT 9  
T30858

glucosyltransferase - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30858  
R:Slimson, C.L., Giffard, P.M., Jacques, N.A.

A:Title: *Streptococcus salivarius* ATCC 25975 possesses at least two genes coding for  
A:Reference number: Z20909; MUID:95122197  
A:Accession: M20909

A:Accession: J50830  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1577 <SIM>  
A:Cross-references: EMBL.L35928; NID.g662380; PID.g662381; PIDN.AAC4143.1.1  
C:Genetics:  
A:Gene: gtfm

Query Match	52.9%	Score 63	DB 2	Length 1577
Best Local Similarity	72.2%	Pred. No. 0.23		
Matches 13, Conservative	0	Mismatches 5	Indels 0	Gaps 0

```

OY      4  HSLLEAWSNDTPYLHD 21
         |||||
Db      591 HSLLEAWSYNHDQYNKD 608

```

RESULT 10  
244911

glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text\_change 15-Oct-1999  
C:Accession: M44811 #S2726: S28809  
R:Giffard, P.M., Simpson, C.L., Milward, C.P., Jacques, N.A.  
J:Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferases  
A:Reference number: M44811; MUID:92148377  
A:Accession: A44811  
A:Molecule type: DNA  
A:Residues: 1-1518 <GI>  
A:Cross-references: EMBL:Z11873; NID:q47526; PIDN:CAA77900.1; PID:q47527  
A:Note: sequences extracted from NCBI backbone (NCBIN:81050, NCBI:P:81052)  
#Genetics:  
A:Gene: gtfJ  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
#:1307-1326/Domain: cpl repeat homology <CP4>

Query Match	48.7%	Score 58;	DB 2;	length 1518;
Best Local Similarity	66.7%	Pred. NO. 1.3;		
Matches 10;	Conservative 2;	Mismatches 3;	Indels 0;	Gaps 0

QY	4	HL\$ILEAWSNDNTPY	18
		: :	
Db	504	HISVLEAWSLNDNHX	518

RESULT 11  
B48445

glyceraldenhyde-3-phosphate dehydrogenase (EC 1.2.1.12) - *Leishmania mexicana*  
C:Species: *Leishmania mexicana*  
C:Date: 31-Dec-1993 #sequence.revision 31-Dec-1993 #text\_change 11-Jun-1999  
C:Accession: B48445; S25142  
R:Hannaert, V.; Blaauw, M.; Kohl, L.; Allert, S.; Opretdoes, F.R.; Michels, P.A.M.  
Mol. Biochem. Parasitol. 55, 115-126, 1992  
A:Title: Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-phosphat  
A:Reference number: A48445; M0UD:91063042  
A:Accession: B48445

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-331 <HAN>  
 A:Cross-references: EMBL:X65220; NID:g9552; PIDN:CAA6323.1; PID:g9553  
 C:Superfamily: glyceroldehyde-3-phosphate dehydrogenase  
 C:Keywords: oxidoreductase

Query Match 47.9%; Score 57; DB 2; Length 331;  
 Best Local Similarity 52.6%; Pred. No. 0.31;  
 Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 NDHLSLEAWSNDTPYLH 20  
 DB 301 NDHFKLVSMYDNETGYSH 319

SULT 12  
 2737

glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius

C:Species: Streptococcus salivarius  
 C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
 C:Accession: S22737; S28810; B4811; S22727

R:Jacques, N.

submitted to the EMBL Data Library, March 1992

A:Reference number: S22726

A:Accession: S22737

A:Molecule type: DNA

A:Residues: 1-1599 <JAC>

A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAA77998.1; PID:g47531

A:Experimental source: ATCC 25975

R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes

A:Reference number: A44811; MUID:92148377

A:Accession: S28810

A:Molecule type: DNA

A:Residues: 1-51 <GIF>

A:Cross-references: EMBL:Z11873

C:Genetics:

A:Gene: gtfK

C:Superfamily: cpl repeat homology

C:Keywords: glycosyltransferase; hexosyltransferase

F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 47.9%; Score 57; DB 2; Length 1599;  
 Best Local Similarity 55.6%; Pred. No. 2;  
 Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 HSLSLAWSNDTPYLH 21  
 DB 494 NISLEAWSNDPYVNE 511

RESULT 13

T30289

pristinamycin I synthase 3 - Streptomyces pristinaespiralis

C:Species: Streptomyces pristinaespiralis

C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 01-Dec-2000

C:Accession: T30289

R:de Grey-Lagard, V.A.; Saurin, W.; Thibaut, D.; Gil, P.; Naudin, L.; Crouzet, J.; Bian

submitted to the EMBL Data Library, February 1997

A:Description: Streptogramin B biosynthesis in Streptomyces pristinaespiralis and Strept

A:Reference number: Z20808

A:Accession: T30289

A:Molecule type: preliminary; translated from GB/EMBL/DBJ

A:Status: preliminary; DNA

A:Residues: 1-4848 <DEC>

A:Cross-references: EMBL:Y11548; NID:e1025755; PID:e307539; PIDN:CAA72312.1

C:Genetics:

A:Gene: snbDE

C:Superfamily: acyl carrier protein homology; acetate-CoA ligase homology

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein

F:511-951/Domain: acetate-CoA ligase homology <ACL1>

F:966-1034/Domain: acyl carrier protein homology <ACP1>

F:1563-2024/Domain: acetate-CoA ligase homology #status: atypical <ACL2>

F:2399-2467/Domain: acyl carrier protein homology <ACP2>

F:2995-3427/Domain: acetate-CoA ligase homology <ACP3>

F:3443-3509/Domain: acyl carrier protein homology <ACP3>

F:4043-4492/Domain: acetate-CoA ligase homology <ACP4>

F:4507-4575/Domain: acyl carrier protein homology <ACP4>

F:998,2431,3473/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 43.7%; Score 52; DB 2; Length 4848;  
 Best Local Similarity 47.1%; Pred. No. 44;  
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSLEAWSNDTPYL 17  
 DB 3972 ADDHARVLGWNDDHP 3988

RESULT 14

T48610

hypothetical protein F18022.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48610

R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24493

A:Accession: T48610

A:Molecule type: preliminary

A:Status: preliminary

A:Residues: 1-347 <BEV>

A:Cross-references: EMBL:DB

A:Experimental source: cultivar Columbia; BAC clone F18022

C:Genetics:

A:Map position: 5

A:Introns: 232/3; 296/3

A:Note: F18022.160

Query Match 42.0%; Score 50; DB 2; Length 347;  
 Best Local Similarity 52.9%; Pred. No. 4;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DHLSLEAWSNDTPYL 19  
 DB 200 DYEQIMEAWSMDKCTLYV 216

RESULT 15

T03983

rf2 nuclear restorer protein - maize

N:Alternate names: aldehyde dehydrogenase homolog

C:Species: Zea mays (maize)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000

C:Accession: T03983

R:Cui, X.; Wise, R.P.; Schnable, P.S.

Science 272, 1334-1336, 1996

A:Title: The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize.

A:Reference number: Z15177; MUID:96243131

A:Accession: T03983

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-549 <CUI>

A:Cross-references: EMBL:U43082; NID:g1421729; PIDN:AAC49371.1; PID:g1421730

A:Experimental source: strain B73

C:Genetics:

A:Map position: 9

A:Note: rf2

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:107-370/Domain: aldehyde dehydrogenase homology <ALD>

Query Match: 41.6%; Score 49.5; DB 2; Length 549;  
Best Local Similarity 58.8%; Pred. No. 8.2;  
Matches 10; Conservative 1; Mismatches 5; Indels 1; Gaps 1;  
QY 2 NDHLSILEAWSNDNDTPY 18  
|||:||||||  
Db 147 NDELALETW-DNGKPY 162

Search completed: March 27, 2002, 14:01:18  
Job time: 480 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:02 ; Search time 53.4 Seconds  
(without alignments)  
14.419 Million cell updates/sec

Title: US-09-290-049a-10

Perfect score: 119

Sequence: 1 ANDHLSILEAMSDNDPRLHD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	1476	1	GTFB_STRMU
2	110	92.4	1375	1	GTFB_STRMU
3	108	90.8	1592	1	GTFB_STRMU
4	108	90.8	1597	1	GTFB_STRMU
5	78	65.5	1365	1	GTFB_STRMU
6	75	63.0	1462	1	GTFB_STRMU
7	57	47.9	330	1	GTFB_STRMU
8	49	41.2	336	1	GTFB_STRMU
9	47.5	39.9	330	1	GTFB_STRMU
10	46	38.7	275	1	GTFB_STRMU
11	46	38.7	335	1	GTFB_STRMU
12	46	38.7	476	1	GTFB_STRMU
13	45	37.8	236	1	GTFB_STRMU
14	45	37.8	385	1	GTFB_STRMU
15	44.5	37.4	470	1	GTFB_STRMU
16	44	37.0	234	1	GTFB_STRMU
17	44	37.0	330	1	GTFB_STRMU
18	44	37.0	332	1	GTFB_STRMU
19	44	37.0	332	1	GTFB_STRMU
20	44	37.0	337	1	GTFB_STRMU
21	44	37.0	337	1	GTFB_STRMU
22	44	37.0	337	1	GTFB_STRMU
23	44	37.0	337	1	GTFB_STRMU
24	44	37.0	337	1	GTFB_STRMU
25	44	37.0	337	1	GTFB_STRMU
26	43.5	36.6	115	1	GTFB_STRMU
27	43.5	36.6	130	1	GTFB_STRMU
28	43.5	36.6	130	1	GTFB_STRMU
29	43.5	36.6	130	1	GTFB_STRMU
30	43.5	36.6	130	1	GTFB_STRMU
31	43	36.1	235	1	GTFB_STRMU
32	43	36.1	236	1	GTFB_STRMU
33	43	36.1	320	1	GTFB_STRMU

34	43	6.1	322	1	G3P_SHEEP
35	43	6.1	332	1	G3P_PIG
36	43	36.1	332	1	G3P_RABIT
37	43	36.1	334	1	G3P2_HUMAN
38	43	36.1	337	1	G3P_MYCE
39	43	36.1	424	1	IVD_RAT
40	43	36.1	674	1	COA_BACTJ
41	43	36.1	2485	1	PTND_HUMAN
42	42.5	35.7	441	1	HGD_STRCO
43	42.5	35.7	491	1	G6PD_ECOLI
44	42.5	35.7	645	1	IME2_YEAST
45	42.5	35.7	764	1	YIS7_YEAST

## ALIGNMENTS

RESULT	ID	Query Match	Length	ID	Description
1	GTFB_STRMU	100.0	1476	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
2	GTFB_STRMU	92.4	1375	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
3	GTFB_STRMU	90.8	1592	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
4	GTFB_STRMU	90.8	1597	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
5	GTFB_STRMU	65.5	1365	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
6	GTFB_STRMU	63.0	1462	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
7	GTFB_STRMU	47.9	330	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
8	GTFB_STRMU	41.2	336	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
9	GTFB_STRMU	39.9	330	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
10	GTFB_STRMU	38.7	275	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
11	GTFB_STRMU	38.7	335	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
12	GTFB_STRMU	38.7	476	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
13	GTFB_STRMU	37.8	236	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
14	GTFB_STRMU	37.8	385	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
15	GTFB_STRMU	37.4	470	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
16	GTFB_STRMU	37.0	234	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
17	GTFB_STRMU	37.0	330	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
18	GTFB_STRMU	37.0	332	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
19	GTFB_STRMU	37.0	332	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
20	GTFB_STRMU	37.0	337	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
21	GTFB_STRMU	37.0	337	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
22	GTFB_STRMU	37.0	337	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
23	GTFB_STRMU	37.0	337	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
24	GTFB_STRMU	37.0	337	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
25	GTFB_STRMU	37.0	337	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
26	GTFB_STRMU	36.6	115	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
27	GTFB_STRMU	36.6	130	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
28	GTFB_STRMU	36.6	130	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
29	GTFB_STRMU	36.6	130	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
30	GTFB_STRMU	36.6	130	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
31	GTFB_STRMU	36.1	235	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
32	GTFB_STRMU	36.1	236	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
33	GTFB_STRMU	36.1	320	AA	STRAIN-67308013; PubMed=3040685; MEDLINE=87308013; PubMed=3040685; Shiroza T., Ueda S., Kuramitsu H. K.; "Sequence analysis of the gtfb gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).

```

DR EMBL: D88657; BAA26109.1; -
DR EMBL: D88660; BAA26113.1; -
DR EMBL: D89977; BAA26119.1; -
DR PIR: B3135; B3135.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro.70.
DR Pfam: PF01473; CW_binding_1.13.
DR Pfam: PF02324; Glyco_hydro.70; 1.
KW Transferase: Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1476
FT DOMAIN 35 1051
FT REPEAT 1097 1476
FT REPEAT 1097 1130
FT REPEAT 1161 1470
FT REPEAT 1161 1210
FT REPEAT 1225 1275
FT REPEAT 1290 1340
FT REPEAT 1355 1405
FT REPEAT 1420 1470
FT REPEAT 62 62
FT VARIANT 65 65
FT VARIANT 68 68
FT VARIANT 78 78
FT VARIANT 86 86
FT VARIANT 89 89
FT VARIANT 168 168
FT VARIANT 276 276
FT VARIANT 399 399
FT VARIANT 474 474
FT VARIANT 512 512
FT VARIANT 519 519
FT VARIANT 701 701
FT VARIANT 708 708
FT VARIANT 938 938
FT VARIANT 952 952
FT VARIANT 963 964
FT VARIANT 968 970
FT VARIANT 1086 1086
FT VARIANT 1158 1158
FT VARIANT 1163 1163
FT VARIANT 1168 1168
FT VARIANT 1182 1182
FT VARIANT 1234 1234
FT VARIANT 1263 1263
FT VARIANT 1263 1263
FT VARIANT 1264 1264
FT VARIANT 1272 1272
FT VARIANT 1329 1329
FT VARIANT 1394 1394
FT VARIANT 1402 1402
FT VARIANT 1459 1459
FT CONFLICT 570 570
FT CONFLICT 800 817
FT CONFLICT 1310 1310
FT CONFLICT 1476 AA; 165685 MW; 34798628076544D98 CRC64;
SQ SEQUENCE 1476 AA; 165685 MW; 34798628076544D98 CRC64;

Query Match 100.0%; Score 119; DB 1; Length 1476;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT 2
ID GTFC_STRMU STANDARD; PRT; 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1998 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE=89137980; PubMed=2976010;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the glfC gene from Streptococcus mutans GS-5."
RL Gene 69:101-109(1988).
RN [2]
RP SEQUENCE OF 1-349 FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the glfB gene from Streptococcus mutans."
RL J. Bacteriol. 169:4263-4270(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-SI SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-SI SYNTHESIZES BOTH
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC
DR EMBL: M22054; AAA88592.1; -
DR EMBL: M17361; AAA88589.1; -
DR PIR: J07345; J07345.
DR PIR: C3135; C3135.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro.70.
DR Pfam: PF01473; CW_binding_1.7.
DR Pfam: PF02324; Glyco_hydro.70; 1.
KW Transferase: Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1375
FT DOMAIN 35 1050
FT DOMAIN 1126 1375
FT REPEAT 1126 1375
FT REPEAT 1126 1159
FT REPEAT 1169 1200
FT REPEAT 1227 1238
FT REPEAT 1253 1303
FT REPEAT 1318 1330
FT REPEAT 1375 AA; 153022 MW; D4B80CBBE0AACE13 CRC64;
SQ SEQUENCE 1375 AA; 153022 MW; D4B80CBBE0AACE13 CRC64;

```





FT	REPEAT	1163	1213	AC REPEAT.
FT	REPEAT	1227	1277	AC REPEAT.
FT	REPEAT	1392	1342	AC REPEAT.
FT	REPEAT	1352	1399	B REPEAT.
FT	REPEAT	1406	1455	AC REPEAT.
FT	REPEAT	1465	1512	B REPEAT.
FT	REPEAT	1519	1568	AC REPEAT.
FT	REPEAT	1582	1597	A REPEAT. (INCOMPLETE).
Q	SEQUENCE	1597 AA:	177080 MM:	B9E66A200868798E CRC64;

Query Match	90.8%	Score 108;	DB 1;	Length 1597;
Best Local Similarity	85.7%	Pred. No. 1.2e-08;		
Matches 18; Conservative	3;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 ANDHLSILEAWSNDNTPYLHD 21  
11:11:111111111111  
483 ANNHVSIWEAWSNDNTPYLHD 503

RESULT	5	
GTFS_STRDO		
ID	GTFS_STRDO	STANDARD;
		PRT; 1365 AA

DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE GLUCOSYLTRANSFERASE-S, PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
NN (GTF-S).  
NN (DEXTRANSUCRASE).

OS *Streptococcus downei* (*Streptococcus sobrinus*).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus  
OX NCBI\_TaxID=1317;  
[1]  
RN  
RP  
RC STRAIN=MEE28;  
RX MEDLINE=90316665; PubMed=2142479;  
RA Gilmore K.S., Russell R.R., Ferretti J.J.;  
RT "Analysis of the *Streptococcus downei* gts gene, which specifies a  
RT glucosyltransferase that synthesizes soluble glucans.";  
RL Infect. Immun. 58:2452-2458(1990).

CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).

CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
CC PRIMER GLUCAN UNLIKE GTF-1.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA  
CC 1,6-GLUCOSE).  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN  
CC BINDING PROTEIN FROM S.MUTANS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on  
CC its use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcementc>  
CC or send an email to [license@sib-sb.ch](mailto:license@sib-sb.ch)).

CC -----  
DR EMBL: M30943; AAA26898.1; -.  
DR PIR: A41483; A41483.  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 10.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
FT SIGNAL 1 36 OR 37 (POTENTIAL).  
FT CHAIN 37 1365 GLYCOSYLTRANSFERASE-S.

FT	DOMAIN	37	1050	CATALYTIC (APPROXIMATE).
FT	DOMAIN	1083	1365	GLUCAN-BINDING (APPROXIMATE).
FT	DOMAIN	1083	1365	4.5 X TANDEM REPEATS.
FT	REPEAT	1083	1131	1.
FT	REPEAT	1150	1139	2.
FT	REPEAT	1225	1274	3.
FT	REPEAT	1289	1339	4.
FT	REPEAT	1353	1365	5. (INCOMPLETE).
SO	SEQUENCE	1365 AA;	151590 MM;	167296B5AAE8C476 CRC64;

Query Match	65.5%	Score 78	DB 1	Length 1365
Best Local Similarity	71.4%	Pred. No.	0.00044	
Matches 15	Conservative	1	Mismatches 5	Indels 0
			Gaps	0

QY 1 ANDHLSILEAWSNDNDTPYLHD 21  
| | | | | | | | | : |  
Db 467 AIDHLSILEAWSGNDNDYKVD 487

RESULT	6	
CTFD_STRMU		
ID	CTFD_STRMU	STANDARD:
		PRT: 1462 AA.

DT 01-FEB-1996 (Rel. 33, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-S. PRECIOSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFD.

OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae  
CC Streptococcus.  
OX NCBI\_Taxid=1509;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RX MEDLINE=91100958; PubMed=2146600;  
RA Honda O., Kato C., Kuramitsu H.K.;  
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
RL J. Gen. Microbiol. 136:2099-2105(1990).

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4239, MT4245, MT4251, MT4467, AND MT8148;  
RX MEDLINE=98231643; Pubmed=9570124;  
RA Fujiwara T., Tero Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.  
RA Kimura S., Hamada S.;

RT "Molecular analyses of glucosyltransferase genes among strains of  
R1 *Streptococcus mitis* ",  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.

CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
CC TYPES OF GLUCANS.

--  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
--

```

CC -----
DR EMBL; M29296: AAA26895.1; -
DR EMBL; D88653: BAA26103.1; -
DR EMBL; D88656: BAA26107.1; -
DR EMBL; D88659: BAA26111.1; -
DR EMBL; D88662: BAA26115.1; -
DR EMBL; D89979: BAA26121.1; -
DR InterPro: IPR002479: CW_binding.
DR InterPro: IPR003318: Glyco_hydro_70.
DR Pfam: PF01473: CW_binding_1; 11.
DR Pfam: PF03324: Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1
FT CHAIN ? 1462
FT DOMAIN 1332 1423
FT REPEAT 1232 1295
FT REPEAT 1296 1359
FT REPEAT 1360 1423
FT VARIANT 58 58
FT VARIANT 68 68
FT VARIANT 81 81
FT VARIANT 113 113
FT VARIANT 122 122
FT VARIANT 132 132
FT VARIANT 135 135
FT VARIANT 202 202
FT VARIANT 255 255
FT VARIANT 275 275
FT VARIANT 288 288
FT VARIANT 301 301
FT VARIANT 313 313
FT VARIANT 317 317
FT VARIANT 328 328
FT VARIANT 350 350
FT VARIANT 628 633
FT VARIANT 688 688
FT VARIANT 726 732
FT VARIANT 726 730
FT VARIANT 964 964
FT VARIANT 1019 1019
FT VARIANT 1059 1060
FT VARIANT 1060 1060
FT VARIANT 1080 1080
FT VARIANT 1142 1142
FT VARIANT 1198 1198
FT VARIANT 1220 1220
FT VARIANT 1280 1280
FT VARIANT 1282 1282
FT VARIANT 1290 1290
FT VARIANT 1311 1311
FT VARIANT 1403 1403
FT VARIANT 1425 1425
FT VARIANT 1449 1449
FT CONFLICT 1428 1462
FT SEQUENCE 1462 AA; 163512 MW; 5C6541F0DCB0DF00 CRC64;

```

```

Query Match 63.0%; Score 75; DB 1; Length 1462;
Best Local Similarity 71.4%; Pred. No. 0.0014;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 ANDHLSILEAWSNDNDPYLHD 21
I :|||||I I

```

```

DB 495 AINHLSILEAWSNDNDPYXKD 515
RESULT 7
G3PC_LEIME STANDARD; PRT; 330 AA.
AC 001558;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (EC 1.2.1.12)
DE (GAPDH).
GN GAPC.
OS Leishmania mexicana.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. MEXICANA;
RX MEDLINE=93063042; PubMed=1435864;
RA Hannant V., Blaauw M., Kohl L., Allert S., Oppendoes F.R.,
RA Michels P.A.M.;
RT "Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-
RT phosphate dehydrogenase in Leishmania mexicana."
RL Mol. Biochem. Parasitol. 55:115-126(1992).
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.
CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X65220: CAA46323.1; -
DR PIR: S25142; S25142.
DR PIR: B48445; B48445.
DR HSSP: P06977; 1GAD.
DR InterPro: IPR000173; GAP_DH.
DR Pfam: PF00044; gpdh; 1.
DR PRINTS: PR00078; G3PDHGNASE.
DR PROSITE: PS00071; GAPDH; 1.
KW Glycolysis; Oxidoreductase; NAD.
FT INIT_MET 0
FT BINDING 148 148
FT ACT_SITE 175 175
FT SEQUENCE 330 AA; 35511 MW; ED4B6D8BEA207F1E CRC64;

```

```

Query Match 47.9%; Score 57; DB 1; Length 330;
Best Local Similarity 52.6%; Pred. No. 0.15;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 NDHLSILEAWSNDNDPYLHD 20
DB 300 NDHFVKLVSWYDNETGYSH 318

```

```

RESULT 8
ID UL16_EBV STANDARD; PRT; 336 AA.
AC P03221;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BGLF2 PROTEIN.

```

```

CN      BGIF2.
OS      Epstein-Barr virus (strain B95-8) (Human herpesvirus 4) .
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Gammaherpesvirinae; Lymphocryptovirus.
OX      NCBI_TaxID=103377;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=84270667; PubMed=6087149;
RA      Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA      Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA      Tuffnell P.S., Barrett B.G.;
RT      "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL      Nature 310:207-211(1984) .
RP      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92113548; PubMed=1662696;
RA      Chen M.-R., Hsu T.-Y., Lin S.-W., Chen J.-Y., Yang C.-S.;
RA      "Cloning and characterization of cDNA clones corresponding to
RA      transcripts from the BamHI G region of the Epstein-Barr virus genome
RA      and expression of BGIF2.";
RL      J. Gen. Virol. 72:3047-3055(1991) .
CC      -I- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL16,
CC      HSV-6 ORF1R, EBV-1 46, HCMV UL94, EBV BDLF2, AND VZV 44.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; M0514; AAA45871.1; -
DR      EMBL; V01555; CAA24831.1; -
DR      EMBL; S77132; AAB21113.1; -
DR      PIR; A03784; Q0BE40.
DR      PIR; S33036; S33036.
KW      late protein.
SQ      SEQUENCE 336 AA: 36888 MW: 840937A416D5584C CRC64:

```

```

Query Match Similarity      41.2%   Score 49: DB 1; Length 336;
Best Local Similarity      41.2%   Pred No 2.7;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY      1 ANDHLSTLEAMSNDTP 17
      1 1::11 1::11
      216 AGAHVNIILRGWTEDDSP 232

RESULT 9
G3PC_TRYB      STANDARD:      PRT:      330 AA.
ID  G3PC_TRYB
AC  P10097;
DT  01-MAR-1989 (Rel. 10, Created)
DT  01-MAY-1992 (Rel. 22, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (EC 1.2.1.12)
DE  (GAPDH).
OS  Trypanosoma brucei brucei.
OC  Eukaryota, Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
CX  NCB1_FaxID=5702;
RN  (1)
RP  SEQUENCE FROM N.A.
RC  STRAIN=427;
RX  MEDLINE=91249838; PubMed=2040303;
RA  Michels P.A.M., Marchand M., Kohl L., Allert S., Wierenga R.K.,
RA  Opperdoes F.R.;
RT  "The cytosolic and glycosomal isoenzymes of glyceraldehyde-3-phosphate
RT  dehydrogenase in Trypanosoma brucei have a distant evolutionary
RT  relationship."
RL  Eur. J. Biochem. 198:421-428(1991).
RN  (2)

```

```

RE PRELIMINARY SEQUENCE OF 1-85.
RA MEDLINE-87161817; PubMed-3830153;
RX Missel O., van Beunnen J., Lambelr A.M., van der Meer R.,
RA Operdoes F.R.;
RT "Glyceraldehyde-3-phosphate dehydrogenase from Trypanosoma brucei.
RT Comparison of the glycosomal and cytosolic isoenzymes.";
RL Eur. J. Biochem. 162:501-507(1987).
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
CC + NAD(+) -> 1,3-DIPHOSPHATEGLYCERATE + NADH.
CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See license@isb-sib.ch).
CC
CC EMBL; X53472; CA937568.1; -.
DR PIR; S16091; DEUTGC.
DR HSSP; P06977; 1GAE.
DR InterPro; IPR000173; GAP_DH.
DR Pfam; PF00044; gpdb.1.
DR PRINTS; PR00078; G3PDHGRNAS.
DR PROSITE; PS00071; GAPDH; 1.
DR GlycoLysis; Oxidoreductase; NAD.
FT
FT INT_MET 0
FT BINDING 149 149 GLYCERALDEHYDE 3-PHOSPHATE.
FT ACT_SITE 176 176 ACTIVE SITE THIOL GROUP DURING CATALYSIS.
FT VARIANT 72 72 R -> I.
FT CONFLICT 56 56 A -> T (IN REF. 2).
FT CONFLICT 60 60 Q -> K (IN REF. 2).
SQ SEQUENCE 330 AA; 35503 MW; 7E1FA0FD091E7C5 CRC64;

```

```

Query Match          39.9%:  Score 47.5;  DB 1;  Length 330;
Best Local Similarity 43.5%:  Pred. No. 4.5;
Matches 10;  Conservative 4;  Mismatches 6;  Indels 3;  Gaps 1;

QY      2 NDHLSILEAWSNDNTPY---LHD 21
      11: 1:11111:11
Db      301 NDNFVKLTWSYWDNETGYSNKKVHD 323

RESULT 10
YG33_YEAST
AC      P53271;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      HYPOHETICAL 31.8 KDA PROTEIN IN NUP57-NEP1 INTERGENIC REGION.
GN      YGR120C OR G6324.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;

RM      11
RM      SEQUENCE FROM N.A.
RP      STRAIN=S288C / FY1679;
RX      MEDLINE=97197982; PubMed=9046098;
RT      van Dyck L., Tettelin H., Purnelle B., Goffeau A.;
RT      "An 18.3 kb DNA fragment from yeast chromosome VII carries four
RT      unknown open reading frames, the gene for an Asn synthase, remnants
RT      of Ty and three tRNA genes.";
RL      Yeast 13:171-176(1997).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X83099: CAA58155.1: -  
DR EMBL: 272805: CAA97130.1: -  
DR SGD: S0003352: SEC35.  
KW Hypothetical protein.  
SQ SEQUENCE 275 AA; 31799 MW; AA102D086FE3FAD7 CRC64;

Query Match 38.7%; Score 46; DB 1; Length 275;  
Best Local Similarity 50.0%; Pred. No. 6.1;  
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 7 ILEAMSDNDTPYLH 20  
|||:|::|::|::|  
217 ILESCADSNSPRIH 230

RESULT 11  
G3P\_CHLPN STANDARD; PRT: 335 AA.  
AC 09J270: 09J0H7;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH).  
GN GAP OR GAPA OR CPN0624 OR CP0123.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.  
OX NCBI\_TaxID=83538;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CML029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RL Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA White T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA Read O., Hickey E.K., Peterson J., Ullrich T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Eisen J., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shiba T., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shirai T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CML029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
CC -I- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
CC + NAD(+) = 1,3-DIPHOSPHATE + NADH.  
CC -I- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
CC -I- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC -I- SUBCELLULAR LOCATION: CYTOSOL.  
CC -I- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
CC DEHYDROGENASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE001647: AAD18763.1: -  
DR EMBL: AE002173: AAF38006.1: -  
DR EMBL: AP002547: BAA9831.1: -  
DR HSSP: P06977: 1GAE.  
DR TIGR: CP0123: -  
DR InterPro: IPR000173: GAP\_DH.  
DR Pfam: PF00044: gpdb: 1.  
DR PRINTS: PR00078: G3PDHRCNASE.  
DR PROSITE: PS00071: GAPDH\_1.  
KW Glycolysis; Oxidoreductase; NAD; Complete proteome.  
FT BINDING 151  
FT ACT\_SITE 178 178  
FT ACTIVATES THIOL GROUP DURING CATALYSIS  
FT SEQUENCE 335 AA; 36837 MW; C86DDE3AD3ADFEE CRC64;

Query Match 38.7%; Score 46; DB 1; Length 335;  
Best Local Similarity 52.9%; Pred. No. 7.7;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 2 NDHLSILEAMSDNDTPY 18  
|||:|::|::|::|  
DB 303 NDRFRLVAMWDENERGY 319

RESULT 12  
PCHA\_PSEAE STANDARD; PRT: 476 AA.  
AC 051508;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE SALICYLATE BIOSYNTHESIS ISOCORISMATE SYNTHASE (EC 5.4.99.6).  
GN PCHA OR PA4231.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 15692 / PA01;  
RX MEDLINE=2043737; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Medman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Structural genes for salicylate biosynthesis from chorismate in  
RT Pseudomonas aeruginosa.";  
RL Mol. Gen. Genet. 249:217-228(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 15692 / PA01;  
RX MEDLINE=2043737; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Medman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Structural genes for salicylate biosynthesis from chorismate in  
RT Pseudomonas aeruginosa.";  
RL Nature 406:959-964(2000).  
CC -I- FUNCTION: INVOLVED IN THE CONVERSION OF CHORISMATE TO SALICYLATE  
CC (PROBABLY).  
CC -I- CATALYTIC ACTIVITY: CHORISMATE = ISOCORISMATE.  
CC -I- PATHWAY: SALICYLATE BIOSYNTHESIS.  
CC -I- SIMILARITY: STRONG, TO OTHER ISOCORISMATE SYNTHASES; WEAK, TO  
CC TRE AND PAB.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X82644; CAA57969.1; -  
 DR EMBL: AB04840; AAG07619.1; -  
 DR InterPro: IPR000350; Choriismate\_bind.  
 DR Pfam: PF00425; choriismate\_bind.1.  
 DR Prodom: PD000779; Choriismate\_bind.1.  
 DR Isomerase: Complete proteome.  
 SQ SEQUENCE 476 AA; 52071 MW; D0F6AFD9C9CF9CFS CRC64;  
 -----  
 Query Match 38.7%; Score 46; DB 1; Length 476;  
 Best Local Similarity 44.4%; Pred. No. 12;  
 Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
 -----  
 QY 3 DHSLEANSNDPVLH 20  
 DB 44 DPLQVFGAMDDRPCLY 61  
 -----  
 RESULT 13  
 GALT\_HUMAN STANDARD; PRT; 236 AA.  
 ID GALT\_HUMAN  
 AC Q14353;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE GUANIDINOACETATE N-METHYLTRANSFERASE (EC 2.1.1.2).  
 GN GALT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=96138544; PubMed=8547310;  
 RA Isbrandt D., von Figura K.;  
 RT "Cloning and sequence analysis of human guanidinoacetate N-  
 methyltransferase cDNA.";  
 RT Biochim. Biophys. Acta 1264:265-267 (1995).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97472276; PubMed=9325156;  
 RA Jenne D.E., Olsen A.S., Zimmer M.;  
 RT "The human guanidinoacetate methyltransferase (GALT) gene maps to a  
 RT syntenic region on 19p13.3, homologous to band C of mouse chromosome  
 RT 10, but GALT is not mutated in jittery mice.";  
 RT Biochem. Biophys. Res. Commun. 238:723-727 (1997).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Isbrandt D., Schmidt A.;  
 RT "Gene structure of human guanidinoacetate N-methyltransferase.";  
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,  
 RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,  
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,  
 RA Dangnan L., Poundstone P., Christensen M., Georgescu A., Avila J.,  
 RA Liu S., Altix C., Andreise T., Frankheim M., Amico-Keller G.,  
 RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,  
 RA Kromboller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,  
 RA Kobayashi A., Olsen A.S., Carrano A.V.;  
 RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + GUANIDINOACETATE =

-----  
 CC S-ADENOSYL-L-HOMOCYSTEINE + CREATINE.  
 CC -1- PATHWAY: LAST STEP OF CREATINE BIOSYNTHESIS.  
 CC -1- DISEASE: DEFECTS IN GALT ARE RESPONSIBLE FOR NEUROLOGIC SYNDROMES  
 CC AND MUSCULAR HYPOTONIA. BIOCHEMICALLY IT IS CHARACTERIZED BY LOW  
 CC EXCRETION OF CREATINE, DEFICIENCY OF CREATINE AND CREATINE  
 CC PHOSPHATE, AND SIMULTANEOUS ACCUMULATION OF GUANIDINOACETATE IN  
 CC BRAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z49878; CAA90035.1; -  
 DR EMBL: AF010248; AAD04781.1; -  
 DR EMBL: AF010246; AAD04781.1; JOINED.  
 DR EMBL: AF010247; AAD04781.1; JOINED.  
 DR EMBL: AF188893; AAF01461.1; -  
 DR EMBL: AC005329; AAC27668.1; -  
 DR MIM: 601240; -  
 DR Transferase: Methyltransferase.  
 SQ SEQUENCE 236 AA; 26318 MW; 6B8E845CE56189F5 CRC64;  
 -----  
 Query Match 37.8%; Score 45; DB 1; Length 236;  
 Best Local Similarity 44.0%; Pred. No. 7.3;  
 Matches 11; Conservative 3; Mismatches 3; Indels 8; Gaps 2;  
 -----  
 QY 1 ANDHSIL-----EAMSDNDPVLH 20  
 DB 30 ADTHLRILCKPVMKX---ETPYMH 51  
 -----  
 RESULT 14  
 GUNF\_FUSOX STANDARD; PRT; 385 AA.  
 ID GUNF\_FUSOX  
 AC P46239;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE PUTATIVE ENDOGLUCANASE TYPE F PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-  
 DE GLUCANASE) (CELLULOSE).  
 OS Fusarium oxysporum.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreales; mitosporic Hypocreales; Fusarium.  
 OX NCBI\_TaxID=5507;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95047531; PubMed=7959045;  
 RA Sheppard P.O., Grant F.J., Oort F.J., Sprecher C.A., Foster D.C.,  
 RA Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;  
 RT "The use of conserved cellulase family-specific sequences to clone  
 RT cellulase homologue cDNAs from Fusarium oxysporum.";  
 RT Cellulose 150:163-167 (1994).  
 RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC  
 CC LINKAGES IN CELLULOSE.  
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL  
 CC HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L29380; AAA65588.1; -  
 DR HSP; P56588; IBG4.







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:03 ; Search time 188.53 Seconds

(without alignments)  
16.293 Million cell updates/sec

Title: US-09-290-049a-10  
Perfect score: 119  
Sequence: 1 ANDHLSLEAWSNDNDPYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_17:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_mammal:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

ult No.	Score	Query Match	Length	ID	Description
1	119	100.0	1390	2	069385 streptococc
2	119	100.0	1455	2	069391 streptococc
3	110	92.4	1455	2	069382 streptococc
4	110	92.4	1455	2	069388 streptococc
5	110	92.4	1455	2	069397 streptococc
6	108	90.8	1590	2	055263 streptococc
7	108	90.8	1590	2	059983 streptococc
8	76	65.5	1338	2	09WXJ4 streptococc
9	76	63.9	1016	2	09LCJ7 streptococc
10	74	62.2	1477	2	09L456 streptococc
11	74	62.2	1508	2	052224 streptococc
12	74	62.2	1508	2	09EZH5 streptococc
13	73	61.3	1575	2	09LCH3 streptococc
14	73	61.3	1577	2	054178 streptococc
15	71	59.7	1527	2	09ZAR4 streptococc
16	66	55.5	1449	2	068542 streptococc
17	66	55.5	1449	2	055264 streptococc
18	63	52.9	1577	2	055265 streptococc
19	62	52.1	1512	2	09WXJ5 streptococc

20	60	50.4	2057	2	09RE05	09e05 leuconostoc
21	58	48.7	1518	2	000600	000600 streptococc
22	57	47.9	1599	2	000599	000599 streptococc
23	52	43.7	4848	2	007944	007944 streptomyces
24	51	42.9	435	10	09LU68	09lu68 arabidopsis
25	50	42.0	347	10	09LU92	09lu92 arabidopsis
26	49.5	41.6	549	10	043274	043274 zea mays (m
27	49	41.2	601	2	005566	005566 mycobacteri
28	48	40.3	490	2	006579	006579 mycobacteri
29	47.5	39.9	519	10	09ZUB6	09zub6 arabidopsis
30	47.5	39.9	549	10	09LUR2	09lur2 oryza sativ
31	47.5	39.9	549	10	09FRX7	09frx7 oryza sativ
32	47	39.5	175	10	09LMJ8	09lmj8 arabidopsis
33	47	39.5	378	10	09SZH3	09szh3 arabidopsis
34	47	39.5	524	2	09KS18	09ks18 vibrio chol
35	47	39.5	525	3	059679	059679 schizosacch
36	46	38.7	195	13	09DFN6	09dfn6 gliliichthys
37	46	38.7	2567	5	09UD06	09ud06 plasmodium
38	45.5	38.2	96	2	09JST4	09jst4 neisseria m
39	45	37.8	286	2	09RDG2	09rdg2 streptomyces
40	45	37.8	333	13	09PW5	09pw5 paraliichthys
41	45	37.8	947	10	023136	023136 arabidopsis
42	45	37.8	1159	4	09UE39	09ue39 homo sapien
43	44.5	37.4	210	2	09IOY4	09ioy4 pseudomonas
44	44.5	37.4	499	5	09VJ56	09vj56 drosophila
45	44	37.0	44	4	016768	016768 homo sapien

## ALIGNMENTS

## RESULT 1

ID 069385 PRELIMINARY: PRT: 1390 AA.

AC 069385;  
DT 01-AUG-1998 (TREMBLREL. 07, Created)  
DT 01-AUG-1998 (TREMBLREL. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBLREL. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-ST.  
GN GTRC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4245;  
RX MEDLINE=9827643; PubMed=9570124;  
RA Fujiwara T., Ierao Y., Hoshino T., Kawabata S., Oshima T., Sobue S.,  
RA Kimura S., Yamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT Streptococcus mutans."  
RL EMBL=9827643; BAA26106.1;  
DR EMBL; D88655; BAA26106.1;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 7.  
DR Pfam; PF03324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1390 AA; 155375 MW; 8847EA956EF05E9F CRC64;

Query Match 100.0%; Score 119; DB 2; Length 1390;  
Best Local Similarity 100.0%; Pred. No. 7.5e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSLEAWSNDNDPYLHD 21  
|||||  
Db 507 ANDHLSLEAWSNDNDPYLHD 527  
RESULT 2  
069391

ID 069391 PRELIMINARY; PRT: 1455 AA.  
AC 069391:  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTRC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4251;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Teruo Y., Hoshino T., Kawabata S., Sobue S.,  
Kimura S., Hamada S.;  
RA "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans."  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88661; BAA26114.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 100.0%; Score 119; DB 2; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 7.9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDTPYLHD 21  
Db 507 ANDHLSILEAMSDNDTPYLHD 527

RESULT 3  
ID 069382 PRELIMINARY; PRT: 1455 AA.  
AC 069382:  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTRC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT8148;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Teruo Y., Hoshino T., Kawabata S., Sobue S.,  
Kimura S., Hamada S.;  
RA "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans."  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88652; BAA26102.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162969 MW; 27DAD3A1ECCA2939 CRC64;

Query Match 92.4%; Score 110; DB 2; Length 1455;  
Best Local Similarity 95.2%; Pred. No. 2e-08;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDTPYLHD 21  
Db 507 ANDHLSILEAMSDNDTPYLHD 527

RESULT 4  
ID 069388 PRELIMINARY; PRT: 1455 AA.  
AC 069388:  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTRC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4239;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Teruo Y., Hoshino T., Kawabata S., Sobue S.,  
Kimura S., Hamada S.;  
RA "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans."  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88658; BAA26110.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 10.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match 92.4%; Score 110; DB 2; Length 1455;  
Best Local Similarity 95.2%; Pred. No. 2e-08;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDTPYLHD 21  
Db 507 ANDHLSILEAMSDNDTPYLHD 527

RESULT 5  
ID 069397 PRELIMINARY; PRT: 1455 AA.  
AC 069397:  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTRC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4467;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Teruo Y., Hoshino T., Kawabata S., Sobue S.,  
Kimura S., Hamada S.;  
RA "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans."  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D89978; BAA26120.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.

SQ SEQUENCE 1455 AA; 162913 MW; A1263427BF24EBE1 CRC64;

Query Match 92.4%; Score 110; DB 2; Length 1455;

Best Local Similarity 93.2%; Pred. No. 2e-08; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDPTPLHD 21  
|||||  
DB 507 ANDHLSILEAMSYNDPTPLHD 527

RESULT 6

ID 055263 PRELIMINARY; PRT; 1590 AA.

AC 055263;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE GTF-I.

OS Streptococcus sobrinus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OX NCBI\_TaxID=1310;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 33478;

RA Sato S.;

RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).

DR EMBL; D63570; BAA09792.1; -

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR00318; Glyco\_hydro\_70.

DR Pfam: PF01473; CW\_binding\_1; 15

DR Pfam: PF02324; Glyco\_hydro\_70; 1.

DR Transferase.

KM SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 90.8%; Score 108; DB 2; Length 1590;

Best Local Similarity 85.7%; Pred. No. 4.5e-08; Mismatches 3; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDPTPLHD 21  
|||  
DB 477 ANNHVSIVEAMSDNDPTPLHD 497

RESULT 7

ID 059983 PRELIMINARY; PRT; 1590 AA.

AC 059983;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE

6-GLUCOSYLTRANSFERASE).

GN GTFI.

OS Streptococcus sobrinus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OX NCBI\_TaxID=1310;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OM176;

RA MEDLINE=94146405; PubMed=8312602;

RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;

RT "DNA sequence of the glucosyltransferase gene of serotype d

RT Streptococcus sobrinus.";

RL DNA Seq. 4:19-27(1993).

CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) - D-

CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).

DR EMBL; D13858; BAA02976.1; -

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR00318; Glyco\_hydro\_70.

DR Pfam: PF01473; CW\_binding\_1; 16.

DR Pfam: PF02324; Glyco\_hydro\_70; 1.

KM Signal; Transferase; Glycosyltransferase.

FT SIGNAL 1 38 POTENTIAL.

FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-1.

SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match 90.8%; Score 108; DB 2; Length 1590;

Best Local Similarity 85.7%; Pred. No. 4.5e-08; Mismatches 3; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDPTPLHD 21  
|||  
DB 477 ANNHVSIVEAMSDNDPTPLHD 497

RESULT 8

ID 09WXJ4 PRELIMINARY; PRT; 1338 AA.

AC 09WXJ4;

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GTF-S.

GN GTF-S.

OS Streptococcus criceti.

OC Plasmid PAM1.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OX NCBI\_TaxID=1333;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HS-6;

RA Inoue M., Fukui K., Miyagi A.;

RT "S. cricetus glucosyltransferase(gtfS and gtfT) genes.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB026123; BAA77236.1; -

DR HSSP; P06278; IV05.

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR00318; Glyco\_hydro\_70.

DR Pfam: PF01473; CW\_binding\_1; 10.

DR Pfam: PF02324; Glyco\_hydro\_70; 1.

DR Plasmid.

SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 65.5%; Score 78; DB 2; Length 1338;

Best Local Similarity 71.4%; Pred. No. 0.0017; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDPTPLHD 21  
|  
DB 437 AIDHLSILEAMSGNDNDYKVD 457

RESULT 9

ID 09LCJ7 PRELIMINARY; PRT; 1016 AA.

AC 09LCJ7;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE DEXTRANSUCRASE.

GN DSRT.

OS Leucostoc mesenteroides.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

OX NCBI\_TaxID=1245;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL B-512F;



OX NCBI\_TaxID=1303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC10557;  
 RX MEDLINE=20231779; PubMed=10768934;  
 RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;  
 RT "Purification, characterization, and molecular analysis of the gene  
 encoding glucosyltransferase from *Streptococcus oralis*.";  
 RL Infect. Immun. 68:2475-2483(2000).  
 DR EMBL: AB025228; BAA95201.1; -;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 17.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KM Transferase.  
 SQ SEQUENCE 1575 AA; 176792 MW; 772A26E4D7CE543 CRC64;

Query Match 61.3%; Score 73; DB 2; Length 1575;  
 Best Local Similarity 77.8%; Pred. No. 0.012;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Db 4 HLSTLEAWSNDNDPYLHD 21  
 |||||||||  
 549 HLSTLEAWSNDNDPYLHD 566

RESULT 14  
 ID 054178 PRELIMINARY; PRT; 1577 AA.  
 AC 054178; 054247;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE.  
 GN GTFG.  
 OS *Streptococcus gordonii* Challis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 CC Streptococcus.  
 OX NCBI\_TaxID=29390;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CHALLIS;  
 RX MEDLINE=96157084; PubMed=8586195;  
 RA Vickerman M.M., Sulavik M.C., Clewell D.B.;  
 RT "Molecular analysis of *Streptococcus gordonii* glucosyltransferase  
 phase variants.";  
 RL Dev. Biol. Stand. 85:309-314(1995).  
 [2]  
 SEQUENCE OF 1-96 FROM N.A.  
 STRAIN-CHALLIS;  
 MEDLINE=92276337; PubMed=1534326;  
 RA Sulavik M.C., Tardif G., Clewell D.B.;  
 RT "Identification of a gene, rgg, which regulates expression of  
 glucosyltransferase and influences the Spp phenotype of *Streptococcus*  
*gordonii* Challis.";  
 RL J. Bacteriol. 174:3577-3586(1992).  
 DR EMBL: U12643; AAC43483.1; -;  
 DR EMBL: M89776; AAC26969.1; -;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 18.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KM Transferase.  
 SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match 61.3%; Score 73; DB 2; Length 1577;  
 Best Local Similarity 77.8%; Pred. No. 0.012;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 4 HLSTLEAWSNDNDPYLHD 21  
 |||||||||

Db 551 HLSTLEAWSNDNDPYLHD 568

RESULT 15  
 ID 092AR4 PRELIMINARY; PRT; 1527 AA.  
 AC 092AR4;  
 DT 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 GN DEX.  
 OS *Leuconostoc mesenteroides*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 CC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B-512-F;  
 RA Bhatnagar R., Singh D.K.S.;  
 RT "Cloning and Molecular Characterization of Dextranucrase Gene from  
*Leuconostoc mesenteroides* NRRL B-512F.";  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U81374; AAD10952.1; -;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 16.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 59.7%; Score 71; DB 2; Length 1527;  
 Best Local Similarity 66.7%; Pred. No. 0.024;  
 Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 ANDHLSTLEAWSNDNDPYLHD 21  
 |||||||  
 Db 581 ANQHLSTLEAWSNDNDPYLHD 601

Search completed: March 27, 2002, 14:26:04  
 Job time: 1676 sec





```
RESULT 2
US-09-740-274-2
; Sequence 2, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: NICHOLS, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2
```

```
Query Match          100.0%; Score 95; DB 21; Length 1475;
Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 VPSYSFIRAHSEVODLI 19  
|||||

Db 552 VPSYSFIRAHSEVODLIA 570

```
RESULT 3
US-09-290-049-15
; Sequence 15, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01P2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049-15
```

```
Query Match          95.8%; Score 91; DB 16; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 VPSYSFIRAHSEVODLI 18  
|||||

Db 1 VPSYSFIRAHSEVODLI 18

```
RESULT 4
US-09-740-274-4
; Sequence 4, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: NICHOLS, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4
```

```
Query Match          95.8%; Score 91; DB 21; Length 1375;
Best Local Similarity 100.0%; Pred. No. 2.1e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 VPSYSFIRAHSEVODLI 18  
|||||

Db 578 VPSYSFIRAHSEVODLI 595

```
RESULT 5
US-09-290-049-17
; Sequence 17, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01P2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049-17
```

```
Query Match          90.5%; Score 86; DB 16; Length 22;
Best Local Similarity 94.4%; Pred. No. 1e-07;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy 1 VPSYSFIRAHSEVODLI 18  
|||||

Db 1 VPSYSFIRAHSEVODLI 18



RESULT 6  
US-09-290-049-2  
; Sequence 2, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARRIES  
; FILE REFERENCE: FDC98-01P2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: HDS peptide  
US-09-290-049-2

Query Match 88.9%; Score 84.5; DB 16; Length 20;  
Best Local Similarity 95.0%; Pred. No. 1.6e-07;  
Matches 19; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 VPSYSFIRAHDSVODLIA 19  
|||||  
DB 1 VPSYSFIRAHDSVODLIA 20

RESULT 7  
US-09-290-049-19  
; Sequence 19, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARRIES  
; FILE REFERENCE: FDC98-01P2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 19  
LENGTH: 22  
TYPE: PRT  
ORGANISM: S. sobrinus  
US-09-290-049-19

Query Match 88.4%; Score 84; DB 16; Length 22;  
Best Local Similarity 88.9%; Pred. No. 2.3e-07;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVODLI 18  
|||||  
DB 1 VPSYSFIRAHDSVODLI 18

RESULT 8  
US-09-290-049-18  
; Sequence 18, Application US/09290049

; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARRIES  
; FILE REFERENCE: FDC98-01P2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 18  
LENGTH: 22  
TYPE: PRT  
ORGANISM: S. downei  
US-09-290-049-18

Query Match 76.8%; Score 73; DB 16; Length 22;  
Best Local Similarity 78.9%; Pred. No. 2.1e-05;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVODLIA 19  
||:|  
DB 1 VPNYVFIRAHDSVOTRIA 19

RESULT 9  
US-09-499-203-2  
; Sequence 2, Application US/09499203  
; GENERAL INFORMATION:  
; APPLICANT: KOSSMANN, Jens  
; APPLICANT: WELSH, Thomas  
; APPLICANT: GUANZ, Martin  
; APPLICANT: KNUTH, Karola  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
; FILE REFERENCE: 147-196P  
; CURRENT APPLICATION NUMBER: US/09/499,203  
; CURRENT FILING DATE: 2000-02-08  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: Patentlin Ver. 2.1  
; SEQ ID NO 2  
LENGTH: 2057  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-499-203-2

Query Match 69.5%; Score 66; DB 18; Length 2057;  
Best Local Similarity 61.1%; Pred. No. 0.099;  
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVODLI 18  
:|:|:|:|:|:|:|:|  
DB 757 IPNYSFVRADHYDAQDPI 774

RESULT 10  
US-09-290-049-16  
; Sequence 16, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARRIES  
; FILE REFERENCE: FDC98-01P2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13

EARLIER APPLICATION NUMBER: 60/115,142  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 16  
LENGTH: 22  
TYPE: PRT  
ORGANISM: S. mutans  
US-09-290-049-16

Query Match 68.4%; Score 65; DB 16; Length 22;  
Best Local Similarity 76.5%; Pred. No. 0.00055;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 SYSFIRAHDESEVODLIA 19  
:|||||||||:11  
3 NYIFIRAHDESEVOTVIA 19

RESULT 11  
US-09-649-885-2  
Sequence 2, Application US/09649885  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starch and  
TITLE OF INVENTION: Latexes in Paper Manufacture  
FILE REFERENCE: 0358D2  
CURRENT APPLICATION NUMBER: US/09/649,885  
CURRENT FILING DATE: 2000-08-28  
PRIOR APPLICATION NUMBER: US 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 2  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-649-885-2

Query Match 68.4%; Score 65; DB 20; Length 1430;  
Best Local Similarity 76.5%; Pred. No. 0.095;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

3 SYSFIRAHDESEVODLIA 19  
:|||||||||:11  
Db 576 NYIFIRAHDESEVOTVIA 592

RESULT 12  
US-09-740-274-6  
Sequence 6, Application US/09740274  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Glucan-containing Compositions and Paper  
FILE REFERENCE: 0357CRD  
CURRENT APPLICATION NUMBER: US/09/740,274  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 09/210,361  
PRIOR FILING DATE: 1998-12-11  
PRIOR APPLICATION NUMBER: 09/007,999  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16

PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 6  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 68.4%; Score 65; DB 21; Length 1430;  
Best Local Similarity 76.5%; Pred. No. 0.095;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 SYSFIRAHDESEVODLIA 19  
:|||||||||:11  
Db 576 NYIFIRAHDESEVOTVIA 592

RESULT 13  
US-09-733-089-19824  
Sequence 19824, Application US/09733089  
GENERAL INFORMATION:  
APPLICANT: Dotson, Stanton B.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: Lutfiyya, Linda L.  
APPLICANT: McIninch, James  
APPLICANT: Wu, Wei  
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
TITLE OF INVENTION: Transcription In Plants  
FILE REFERENCE: 38-21(15300)D  
CURRENT APPLICATION NUMBER: US/09/733,089  
CURRENT FILING DATE: 2000-12-11  
PRIOR APPLICATION NUMBER: US 09/474,435  
PRIOR FILING DATE: 1999-12-28  
PRIOR APPLICATION NUMBER: US 09/654,617  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: US 09/620,392  
PRIOR FILING DATE: 2000-07-19  
NUMBER OF SEQ ID NOS: 24143  
SEQ ID NO: 19824  
LENGTH: 263  
TYPE: PRT  
ORGANISM: Oryza sativa  
US-09-733-089-19824

Query Match 52.1%; Score 49.5; DB 21; Length 263;  
Best Local Similarity 45.0%; Pred. No. 6.7;  
Matches 9; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 VPSYSFIRAHDS-EVODLIA 19  
:|||||:11  
Db 116 LPSFYPVRAHNGEISDTAA 135

RESULT 14  
US-09-816-660-19824  
Sequence 19824, Application US/09816660  
GENERAL INFORMATION:  
APPLICANT: Dotson, Stanton B.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: Lutfiyya, Linda L.  
APPLICANT: McIninch, James  
APPLICANT: Wu, Wei  
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
TITLE OF INVENTION: Transcription In Plants  
FILE REFERENCE: 38-21(15300)D  
CURRENT APPLICATION NUMBER: US/09/816,660  
CURRENT FILING DATE: 2001-03-26

; PRIOR APPLICATION NUMBER: US 09/474,435  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: US 09/654,617  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: US 09/733,089  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: ) US 09/684,016  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: US 09/620,392  
; PRIOR FILING DATE: 2000-07-19  
; NUMBER OF SEQ ID NOS: 24143  
; SEQ ID NO 19624  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-09-816-660-19824

/ Query Match 52.1%; Score 49.5; DB 22; Length 263;  
Best Local Similarity 45.0%; Pred. No. 6.7;  
Matches 9; Conservative 6; Mismatches 4; Indels 1; Gaps 1;  
1 VPSYSFIRAHDS-EVODLIA 19  
:11: :1111: 1: 1 1  
Db 116 LPSEFYVRAHDNGEISDTAA 135

RESULT 15  
US-09-417-507-26458  
; Sequence 26458, Application US/09417507  
; GENERAL INFORMATION:  
; APPLICANT: KEITH G. WEINSTOCK ET AL.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS  
; FILE REFERENCE: PATH99-10  
; CURRENT APPLICATION NUMBER: US/09/417,507  
; CURRENT FILING DATE: 1999-10-14  
; NUMBER OF SEQ ID NOS: 44312  
; SEQ ID NO 26458  
; LENGTH: 541  
; TYPE: PRT  
; ORGANISM: A.fumigatus  
US-09-417-507-26458

Query Match 49.5%; Score 47; DB 18; Length 541;  
Best Local Similarity 47.6%; Pred. No. 46;  
Matches 10; Conservative 4; Mismatches 5; Indels 2; Gaps 1;  
1 VPSYSFIRAHDS-EVODLIA 19  
11:11: 11: 11:11  
204 VPOHSFLAHDQDQASVOSLLA 224

Search completed: March 27, 2002, 14:20:23  
Job time: 1570 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:45 ; Search time 137.48 Seconds

(without alignments)  
9.536 Million cell updates/sec

Title: US-09-290-049A-2

Perfect score: 95

Sequence: 1 VPSYSTRAHDSVQDLIA 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 6899538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCr\_NEW.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW.COMB.pep1:\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW.COMB.pep1:\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95	100.0	19	6	US-09-562-328-27
2	95	100.0	19	6	US-09-290-049A-2
3	95	100.0	23	6	US-09-562-328-44
4	91	95.8	22	6	US-09-562-328-28
5	91	95.8	22	6	US-09-290-049A-15
6	86	90.5	22	6	US-09-562-328-30
7	86	90.5	22	6	US-09-290-049A-17
8	84.5	88.9	20	6	US-09-562-328-26
9	84	88.4	22	6	US-09-562-328-32
10	84	88.4	22	6	US-09-290-049A-19
11	80	84.2	523	6	US-09-604-957-5
12	73	76.8	22	6	US-09-562-328-31
13	73	76.8	22	6	US-09-290-049A-18
14	67	70.5	584	6	US-09-604-957-6
15	66	69.5	535	6	US-09-604-957-7
16	66	69.5	1278	6	US-09-604-957-3
17	65	68.4	22	6	US-09-562-328-29
18	65	68.4	22	6	US-09-290-049A-16
19	65	68.4	545	6	US-09-604-957-4
20	47	49.5	274	6	US-09-675-784A-9123
21	44	46.3	327	8	US-60-356-051-2881
22	41	43.2	633	6	US-09-605-703B-440
23	40.5	42.6	389	6	US-09-708-427-10452
24	40.5	42.6	398	6	US-09-708-427-10451
25	40.5	42.6	414	6	US-09-708-427-10450

26	40.5	42.6	445	6	US-09-614-150-4236	Sequence 4236, Ap
27	40	42.1	547	6	US-09-943-857-2	Sequence 2, Appl
28	40	42.1	547	6	US-09-943-857-4	Sequence 4, Appl
29	40	42.1	547	6	US-09-943-857-6	Sequence 6, Appl
30	40	42.1	572	1	PCR-US02-03987-5626	Sequence 5626, Ap
31	40	42.1	572	6	US-09-815-242-5626	Sequence 5626, Ap
32	40	42.1	572	6	US-10-072-851-5626	Sequence 5626, Ap
33	40	42.1	573	1	PCR-US02-03987-12247	Sequence 12247, A
34	40	42.1	573	1	PCR-US02-03987-12921	Sequence 12921, A
35	40	42.1	573	6	US-09-815-242-12247	Sequence 12247, A
36	40	42.1	573	6	US-09-815-242-12921	Sequence 12921, A
37	40	42.1	573	7	US-10-072-851-12247	Sequence 12247, A
38	40	42.1	573	7	US-10-072-851-12921	Sequence 12921, A
39	40	42.1	737	6	US-09-708-427-2859	Sequence 2859, Ap
40	40	42.1	760	6	US-09-708-427-2858	Sequence 2858, Ap
41	40	42.1	783	6	US-09-708-427-2857	Sequence 2857, Ap
42	39	41.1	242	6	US-09-675-784A-9068	Sequence 9068, Ap
43	39	41.1	309	6	US-09-609-360C-37	Sequence 37, Appl
44	39	41.1	309	6	US-09-345-473E-37	Sequence 37, Appl
45	39	41.1	361	6	US-09-605-703B-1852	Sequence 1852, Ap

## ALIGNMENTS

Result 1  
US-09-562-328-27  
Sequence 27, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES  
FILE REFERENCE: 04995, 0046-01  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288, 965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 27  
LENGTH: 19  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-27

Query Match 100.0%; Score 95; DB 6; Length 19;  
Best Local Similarity 100.0%; Pred. No. 5.5e+10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 VPSYSTRAHDSVQDLIA 19  
Db 1 VPSYSTRAHDSVQDLIA 19  
RESULT 2  
US-09-290-049A-2  
Sequence 2, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL CARRIES  
FILE REFERENCE: 1564, 1008-002  
CURRENT APPLICATION NUMBER: US/09/290, 049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081, 550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115, 142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2  
LENGTH: 19  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: HDS peptide  
US-09-290-049A-2

Query Match 100.0%; Score 95; DB 6; Length 19;  
Best Local Similarity 100.0%; Pred. No. 5.5e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVEVDLIA 19  
Db 1 VPSYSFIRAHDSVEVDLIA 19

RESULT 3  
US-09-562-328-44  
Sequence 44, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 44  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-44

Query Match 100.0%; Score 95; DB 6; Length 23;  
Best Local Similarity 100.0%; Pred. No. 7e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVEVDLIA 19  
5 VPSYSFIRAHDSVEVDLIA 23

RESULT 4  
US-09-562-328-28  
Sequence 28, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 28  
LENGTH: 22  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-28

Query Match 95.8%; Score 91; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 3.2e-09;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VPSYSFIRAHDSVEVDLI 18  
Db 1 VPSYSFIRAHDSVEVDLI 18

RESULT 5  
US-09-290-049A-15  
Sequence 15, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL CARIES  
FILE REFERENCE: 1564.1008-002  
CURRENT APPLICATION NUMBER: US/09/290,049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15  
LENGTH: 22  
TYPE: PRT  
ORGANISM: S. mutans  
US-09-290-049A-15

Query Match 95.8%; Score 91; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 3.2e-09;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVEVDLI 18  
Db 1 VPSYSFIRAHDSVEVDLI 18

RESULT 6  
US-09-562-328-30  
Sequence 30, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 30  
LENGTH: 22  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-30

Query Match 90.5%; Score 86; DB 6; Length 22;  
Best Local Similarity 94.4%; Pred. No. 2.4e-08;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVEVDLI 18  
Db 1 VPSYSFIRAHDSVEVDLI 18

RESULT 7  
US-09-290-049A-17

```
; Sequence 17, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. downei
; US-09-290-049a-17
```

```
Query Match          90.5%; Score 86; DB 6; Length 22;
Best Local Similarity 94.4%; Pred. No. 2.4e-08;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 VPSYSFIRAHDESEVODLI 18
Db 1 VPSYSFIRAHDESEVODLI 18
```

```
RESULT 8
US-09-562-328-26
; Sequence 26, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEEB, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Streptococcus sp.
; US-562-328-26
```

```
Query Match          88.9%; Score 84.5; DB 6; Length 20;
Best Local Similarity 95.0%; Pred. No. 3.8e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
```

```
QY 1 VPSYSFIRAHDESEVODLI 19
Db 1 VPSYSFIRAHDESEVODLI 20
```

```
RESULT 9
US-09-562-328-32
; Sequence 32, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEEB, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
```

```
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Streptococcus sp.
; US-09-562-328-32
```

```
Query Match          88.4%; Score 84; DB 6; Length 22;
Best Local Similarity 88.9%; Pred. No. 5.2e-08;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 VPSYSFIRAHDESEVODLI 18
Db 1 VPSYSFIRAHDESEVODLI 18
```

```
RESULT 10
US-09-290-049a-19
; Sequence 19, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. sobrinus
; US-09-290-049a-19
```

```
Query Match          88.4%; Score 84; DB 6; Length 22;
Best Local Similarity 88.9%; Pred. No. 5.2e-08;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 VPSYSFIRAHDESEVODLI 18
Db 1 VPSYSFIRAHDESEVODLI 18
```

```
RESULT 11
US-09-604-957-5
; Sequence 5, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LOBBERT
; APPLICANT: RAHMOUNI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:15 ; Search time 102.51 Seconds  
(without alignments)  
14.119 Million cell updates/sec

Title: US-09-290-049a-2

Sequence: 1 VPSYSTRAHDSVODLIA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95	100.0	1475	2	B3135
2	91	95.8	1375	2	JT0345
3	84	88.4	1592	2	A38175
4	80	84.2	1508	2	T31098
5	79	83.2	1290	2	JC5473
6	73	76.8	1365	2	A41483
7	65	68.4	1431	2	A45866
8	65	68.4	1449	2	T30857
9	65	68.4	1449	2	T30552
10	64	67.4	1518	2	A44811
11	64	67.4	1599	2	S22737
12	61	64.2	1577	2	T30858
13	46	48.4	575	2	S46329
14	44	46.3	327	2	D86474
15	43	45.3	51	2	G82455
16	43	45.3	214	1	RGBSCA
17	43	45.3	1260	2	A86323
18	42	44.2	431	2	S52583
19	42	44.2	563	2	F81888
20	42	44.2	563	2	DB1134
21	42	44.2	644	2	A64877
22	42	44.2	644	2	B85758
23	42	44.2	659	2	A64139
24	42	44.2	932	2	T40216
25	41.5	43.7	713	2	B64743
26	41.5	43.7	713	2	D85503
27	41	43.2	145	2	G69098
28	41	43.2	344	2	JH0511
29	41	43.2	597	2	F82935

30	41	43.2	1146	2	E70204	hypothetical prote
31	40.5	42.6	414	2	T06120	hypothetical prote
32	40	42.1	359	2	G82197	RstA1/RstA2 protei
33	40	42.1	431	2	B37802	crx protein - Erw
34	40	42.1	534	2	S41735	cholesterol estera
35	40	42.1	548	2	S32615	triacylglycerol 11
36	40	42.1	549	2	JN0552	triacylglycerol 11
37	40	42.1	549	2	JN0551	triacylglycerol 11
38	40	42.1	572	2	B28474	phosphotransferase
39	40	42.1	783	2	T00782	probable anthranil
40	40	42.1	895	2	A86410	protein F3M18.22 f
41	40	42.1	1029	2	F86359	hypothetical prote
42	40	42.1	6359	2	T31679	bacitracin synthet
43	39.5	41.6	192	1	UCPEPP	cuticle protein pr
44	39	41.1	145	2	H84174	hypothetical prote
45	39	41.1	145	2	S07957	hypothetical prote

## ALIGNMENTS

RESULT 1  
B3135  
gtfB protein precursor - Streptococcus mutans  
C:Species: Streptococcus mutans  
C:Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999\*  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A:Reference number: A33135; MUID:87308013  
A:Accession: B33135  
A:Status: preliminary  
A:Residues: 1-1475 <SHL>  
A:Molecule type: DNA  
A:Cross-references: GB:M17361; NID:G153639; PIDN:AAA8588.1; PID:G153640  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
submitted to the Protein Sequence Database, September 1990  
A:Reference number: A33128  
A:Accession: A33128  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-171,173-641,'N',643-1475 <SH2>  
A:Experimental source: strain GS-5  
A:Superfamily: cpl repeat homology  
F:1096-1115/Domain: cpl repeat homology <CP1>  
F:1224-1243/Domain: cpl repeat homology <CP2>  
F:1289-1308/Domain: cpl repeat homology <CP3>  
F:1354-1373/Domain: cpl repeat homology <CP4>  
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 100.0%; Score 95; DB 2; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSTRAHDSVODLIA 19  
DB 552 VPSYSTRAHDSVODLIA 570  
RESULT 2  
JT0345  
dextranase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
N:Alternate names: sucrose 6-glucosyltransferase  
C:Species: Streptococcus mutans  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
A:Accession: JT0345; C33135  
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
Gene 69, 101-109, 1988  
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.  
A:Reference number: JT0345; MUID:89137980  
A:Accession: JT0345

A:Molecule type: DNA  
A:Residues: 1-1375 <UED>  
A:Experimental source: GS-5  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 159, 4263-4270, 1997  
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A:Reference number: A31135; MUID:87308013  
A:Accession: C31135  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-349 <SHI>  
A:Cross-references: GB:M17361  
C:Genetics:  
A:Gene: gtfC  
C:Function:  
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
C:Superfamily: cpl repeat homology  
C:Keywords: duplication; glycosyltransferase; hexosyltransferase  
F:35-1375/Product: signal sequence #status predicted <Sig>  
F:1126-1145/Domain: glycosyltransferase #status predicted <Mat>  
F:1126-1145/Domain: cpl repeat homology <CP1>  
F:1253-1372/Domain: cpl repeat homology <CP2>  
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 95.8%; Score 91; DB 2; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPYSFIRAHDSFVODLI 18  
|||||  
Db 578 VPYSFIRAHDSFVODLI 595

RESULT 3  
A8175  
glucosyltransferase precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C>Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
C:Accession: A38175  
R:Abdo, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus  
A:Reference number: A38175; MUID:91123227  
A:Accession: A38175  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1592 <ABO>  
A:Cross-references: GB:D90213; NID:9217032; PIDN:BAAI4241.1; PID:dl014946; PID:9217033  
C:Superfamily: cpl repeat homology  
F:1093-1112/Domain: cpl repeat homology <CP1>  
F:1223-1241/Domain: cpl repeat homology <CP2>  
F:1287-1306/Domain: cpl repeat homology <CP3>  
F:1330-1351/Domain: cpl repeat homology <CP4>  
F:1352-1371/Domain: cpl repeat homology <CP5>  
F:1402-1420/Domain: cpl repeat homology <CP6>  
F:1465-1484/Domain: cpl repeat homology <CP7>  
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 88.4%; Score 84; DB 2; Length 1592;  
Best Local Similarity 88.9%; Pred. No. 8.7e-06;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPYSFIRAHDSFVODLI 18  
|||||  
Db 548 VPYSFIRAHDSFVODII 565

RESULT 4  
T31098  
probable dextransucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides  
C:Species: Leuconostoc mesenteroides

C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C:Accession: T31098  
R:Monchois, V.; Remaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
FEMS Microbiol. Lett. 159, 307-315, 1998  
A:Title: Cloning and sequencing of a gene coding for an extracellular dextransucrase  
A:Reference number: 220981; MUID:98164374  
A:Accession: T31098  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1508 <MON>  
A:Cross-references: EMBL:AF030129; NID:92766611; PID:92766612; PIDN:AAB95453.1  
A:Experimental source: strain NRRL B-1299  
C:Genetics:  
A:Gene: dsrB  
C:Function:  
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 84.2%; Score 80; DB 2; Length 1508;  
Best Local Similarity 73.7%; Pred. No. 3.9e-05;  
Matches 14; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPYSFIRAHDSFVODLIA 19  
|||  
Db 634 IPNVSFVRAHDSFVQIVIA 652

RESULT 5  
JC5473  
dextransucrase (EC 2.4.1.5) - Leuconostoc mesenteroides  
C:Species: Leuconostoc mesenteroides  
C>Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997  
C:Accession: JC5473  
R:Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.  
Gene 182, 23-32, 1996  
A:Title: Cloning and sequencing of a gene coding for a novel dextransucrase from Leuc  
A:Reference number: JC5473; MUID:97136686  
A:Accession: JC5473  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1290 <MON>  
A:Cross-references: GB:U38181  
C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose  
A:Gene: dsrA  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:78-870/Domain: catalytic #status predicted <CAT>  
F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 83.2%; Score 79; DB 2; Length 1290;  
Best Local Similarity 83.3%; Pred. No. 4.9e-05;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 PYSFIRAHDSFVODLIA 19  
|  
Db 388 PYSFIRAHDSFVOTIIA 405

RESULT 6  
A41483  
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C:Accession: A41483  
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltr  
A:Reference number: A41483; MUID:90316665  
A:Accession: A41483  
A:Molecule type: DNA  
A:Residues: 1-1365 <GIL>

Query Match	Score	DB 2;	Length
Best Local Similarity	Pred. NO.	0.014;	
68.48;			1449;
68.88;			

RESULT 11  
S22737  
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: S22737; S28810; B4811; S22727  
R:Jacques, N.

submitted to the EMBL Data Library, March 1992

A:Reference number: S22726

A:Accession: S22737

A:Molecule type: DNA

A:Residues: 1-1599 <JNC>

A:Cross-References: EMBL:211872; NID:947530; PIDN:CAA7898.1; PID:947531

A:Experimental source: ATCC 25975

A:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes

A:Reference number: A44811; MUID:92148377

A:Accession: S28810

A:Molecule type: DNA

A:Residues: 1-51 <GIF>

A:Cross-References: EMBL:211873

C:Genetics:

A:Gene: gltK

Superfamily: cpl repeat homology

Keywords: glycosyltransferase; hexosyltransferase

1456-1475/Domain: cpl repeat homology <CPR>

Query Match

Best Local Similarity 67.4%; Score 64; DB 2; Length 1599;

Matches 12; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 SYSFIRAHSEVODLIA 19

DB 574 TYLFVRAHSEVQTVIA 590

RESULT 12

glucosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius

C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T30858

R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.

Infect. Immun. 63, 609-621, 1995

A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pr

A:Reference number: Z20909; MUID:95122197

A:Accession: T30858

A:Molecule type: DNA

A:Residues: 1-1577 <SIM>

A:Cross-References: EMBL:L35928; NID:9662380; PID:9662381; PIDN:AA041413.1

metics:

A:Gene: gltM

Query Match

Best Local Similarity 64.2%; Score 61; DB 2; Length 1577;

Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 SYSFIRAHSEVODLIA 19

DB 661 NTYFVRAHSEVQTVIA 677

RESULT 13

Intermediate filament protein - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 20-Jun-2000

C:Accession: S46329

R:Dodemont, H.; Riemer, D.; Ledger, N.; Weber, K.

EMBO J. 13, 2625-2638, 1994

A:Title: Eight genes and alternative RNA processing pathways generate an unexpectedly 1a

A:Reference number: S46326; MUID:94283388

A:Accession: S46329

A:Molecule type: DNA

A:Residues: 1-575 <DOD>

A:Status: preliminary

A:Map position: 2

A:Gene: VCA0471

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Cross-References: EMBL:X70833; NID:9312743; PIDN:CAA50181.1; PID:91848062

C:Genetics:

A:Introns: 31/1; 125/3; 220/3; 252/3; 290/3; 348/3; 445/1; 471/3; 563/3

C:Superfamily: Intermediate filament protein Av71

Query Match

Best Local Similarity 48.4%; Score 46; DB 2; Length 575;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 RAHDESEVODLIA 19

DB 248 RHDESEIHDLIA 259

RESULT 14

probable RING zinc finger protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: D86474

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Malli, R.; Marzla

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Salzman, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallio

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: D86474

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-327 <STO>

A:Cross-References: GB:AE005172; NID:911034939; PIDN:AA027096.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match

Best Local Similarity 46.3%; Score 44; DB 2; Length 327;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 YSFIRAHSEVODL 17

DB 185 YPFIRHNDQMDL 198

RESULT 15

Hypothetical protein VCA0471 [imported] - Vibrio cholerae (strain N16961 serogroup O1

C:Species: Vibrio cholerae

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: G82455

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: G82455

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-51 <HEI>

A:Cross-References: GB:AE004379; GB:AE003853; NID:9657865; PIDN:AAF96375.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCA0471

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

A:Map position: 2

Query Match 45.38; Score 43; DB 2; Length 51;  
Best Local Similarity 58.88; Pred. No. 1.6;  
Matches 10; Conservative 4; Mismatches 1; Indels 2; Gaps 1;  
Oy 1 VPSTFIRAHDSFVODL 17  
||:|||||:||||  
Db 15 VPATSAIR--NSEIRDL 29

Search completed: March 27, 2002, 14:01:16  
Job time: 478 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen, Ltd.

OK protein - protein search, using sw model

Run on: March 27, 2002, 13:59:27 ; Search time 87.3 Seconds  
(without alignments)  
5.671 Million cell updates/sec

Title: US-09-290-049a-3

Perfect score: 112

Sequence: 1 TGARTINGQLLYFRANGVYKG 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents-AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PT05.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	112	100.0	1475	3	US-09-007-999-2	Sequence 2, Appli
2	112	100.0	1475	4	US-09-210-361-2	Sequence 2, Appli
3	87	77.7	22	1	US-08-057-162B-3	Sequence 3, Appli
4	78	69.6	1375	4	US-09-210-361-4	Sequence 4, Appli
5	67	59.8	1430	4	US-09-008-172-2	Sequence 2, Appli
6	67	59.8	1430	4	US-09-210-361-6	Sequence 6, Appli
7	66	58.9	1577	2	US-08-793-824-2	Sequence 2, Appli
8	49.5	44.2	811	1	US-08-480-604A-7	Sequence 7, Appli
9	49.5	44.2	811	4	US-08-405-496A-7	Sequence 7, Appli
10	49.5	44.2	811	4	US-08-915-136-7	Sequence 7, Appli
11	49.5	44.2	812	1	US-08-480-604A-29	Sequence 29, Appli
12	49.5	44.2	812	4	US-08-480-604A-6	Sequence 29, Appli
13	49.5	44.2	2710	1	US-08-480-604A-6	Sequence 6, Appli
14	49.5	44.2	2710	2	US-08-405-496A-6	Sequence 6, Appli
15	49.5	44.2	2710	2	US-08-915-136-6	Sequence 6, Appli
16	49.5	44.2	619	1	US-08-465-746-2	Sequence 2, Appli
17	46	41.1	619	1	US-08-214-164-2	Sequence 2, Appli
18	46	41.1	619	2	US-08-467-852A-3	Sequence 3, Appli
19	46	41.1	619	2	US-08-246-636-2	Sequence 2, Appli
20	46	41.1	619	2	US-08-247-491A-3	Sequence 3, Appli
21	46	41.1	619	2	US-08-319-795-2	Sequence 2, Appli
22	46	41.1	619	2	US-08-468-985-2	Sequence 2, Appli
23	46	41.1	619	3	US-08-312-949-2	Sequence 2, Appli
24	46	41.1	641	4	US-08-961-083-160	Sequence 160, App
25	46	41.1	648	1	US-08-072-070-2	Sequence 2, Appli
26	46	41.1	648	1	US-08-469-434-2	Sequence 2, Appli
27	46	41.1	648	1	US-08-214-222-2	Sequence 2, Appli

28	46	41.1	648	2	US-08-467-852A-2	Sequence 2, Appli
29	46	41.1	648	2	US-08-468-718-2	Sequence 2, Appli
30	46	41.1	648	2	US-08-247-491A-2	Sequence 2, Appli
31	46	41.1	648	3	US-08-446-201-3	Sequence 3, Appli
32	46	41.1	695	1	US-08-127-499A-23	Sequence 23, Appli
33	46	41.1	695	1	US-08-482-847-23	Sequence 23, Appli
34	43	38.4	1477	1	US-08-038-682-4	Sequence 4, Appli
35	43	38.4	1477	1	US-08-302-832-4	Sequence 4, Appli
36	43	38.4	1477	2	US-08-530-198-4	Sequence 4, Appli
37	43	38.4	1477	2	US-08-469-880-4	Sequence 4, Appli
38	43	38.4	1477	2	US-08-728-470-4	Sequence 4, Appli
39	43	38.4	1477	2	US-08-617-697-4	Sequence 4, Appli
40	43	38.4	1477	4	US-08-719-641-4	Sequence 4, Appli
41	40	35.7	265	1	US-08-461-731-2	Sequence 2, Appli
42	40	35.7	265	1	US-09-002-072B-2	Sequence 2, Appli
43	40	35.7	265	4	US-09-314-199-2	Sequence 2, Appli
44	40	35.7	265	5	PCT-US94-10465-3	Sequence 3, Appli
45	40	35.7	654	1	US-08-392-828C-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
; US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007, 999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478, 704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-007-999-2

Query Match 100.0%, Score 112, DB 3; Length 1475;
Best Local Similarity 100.0%, Pred. No. 9,3e-11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGARTTTCQLLYFRANGVYKG 22
Db 1300 TGARTTTCQLLYFRANGVYKG 1321

RESULT 2
; US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210, 361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007, 999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478, 704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/009, 620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485, 243
; EARLIER FILING DATE: 1995-06-07
```





```

RESULT 6
US-09-210-361-6
: Sequence 6, Application US/09210361
: Patent No. 6284479
: GENERAL INFORMATION:
: APPLICANT: Nichols, Scott E.
: TITLE OF INVENTION: Substitutes for Modified Starches and
: TITLE OF INVENTION: Latexes in Paper Manufacture
: FILE REFERENCE: 0357CR
: CURRENT APPLICATION NUMBER: US/09/210,361
: CURRENT FILING DATE: 1998-12-11
: EARLIER APPLICATION NUMBER: 09/007,999
: EARLIER FILING DATE: 1998-01-16
: EARLIER APPLICATION NUMBER: 08/478,704
: EARLIER FILING DATE: 1995-06-07
: EARLIER APPLICATION NUMBER: 09/009,620
: EARLIER FILING DATE: 1998-01-20
: EARLIER APPLICATION NUMBER: 08/485,243
: EARLIER FILING DATE: 1995-06-07
: EARLIER APPLICATION NUMBER: 09/008,172
: EARLIER FILING DATE: 1998-01-16
: EARLIER APPLICATION NUMBER: 08/482,711
: EARLIER FILING DATE: 1995-06-07
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 6
: LENGTH: 1430
: TYPE: PRT
: ORGANISM: streptococcus mutans
: US-09-210-361-6

```

```
Matches      13; Conservative      5; Mismatches      4; Indels      0; Gaps      0;
OY          1 TGARTINGOLLYFRANGVQVGK 22
            ||::|| |::| ::| ||||
Db          1332 TGSQTIAKGKLYFASDGKQVKG 1353

RESULT      7
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
APPLICANT: Jacques, Nicholas Anthony
TITLE OF INVENTION: Genetic Manipulation of Plants to
                    Increase Stored Carbohydrates
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Griffith Hack & Co
STREET: Level 8, 168 Walker Street
CITY: No. 5981838th Sydney
STATE: New South Wales
COUNTRY: Australia
ZIP: 2060
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,824
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PM7643
FILING DATE: 24-AUG-1994
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 61 2 9957 5944  
TELEFAX: 61 2 957 6288  
TELEX: 26547  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1577 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus salivarius

[illegible]

RESULT 8  
 US-08-480-604A-7  
 : Sequence 7, Application US/08480604A  
 : Patent No. 5736139  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: KINK, JOHN A.  
 : APPLICANT: THALLEY, BRUCE S.  
 : APPLICANT: PAHVEY, NISHA V.  
 : APPLICANT: FIRCA, JOSEPH R.  
 : APPLICANT: STAFFORD, DOUGLAS C.  
 : TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
 : TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
 : NUMBER OF SEQUENCES: 32  
 : CORRESPONDENCE ADDRESSES:  
 : ADDRESSEE: MEDLEN & CARROLL, LLP  
 : STREET: 220 MONTGOMERY STREET, SUITE 2200  
 : CITY: SAN FRANCISCO  
 : STATE: CALIFORNIA  
 : COUNTRY: UNITED STATES OF AMERICA  
 : ZIP: 94104  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/480,604A  
 : FILING DATE: 07-JUN-1995  
 : CLASSIFICATION: 424  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/422,711  
 : FILING DATE: 14-APR-1995  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/405,496  
 : FILING DATE: 16-MAR-1995  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/329,154  
 : FILING DATE: 25-OCT-1994  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/161,907  
 : FILING DATE: 02-DEC-1993  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 07/985,321  
 : FILING DATE: 04-DEC-1992  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 07/429,791  
 : FILING DATE: 31-OCT-1989  
 : ATTORNEY/AGENT INFORMATION:  
 :



STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-915-136-7

Query Match 44.2%; Score 49.5; DB 4; Length 811;  
Best Local Similarity 52.2%; Pred. No. 2.6;  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

Oy 1 TGARTINGOLLYFRAN-GVQYKG 22  
Db 587 TGLRTIDGKKYFNTTAVATG 609

RESULT 11  
US-08-480-604A-29  
Sequence 29, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,604A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/422,711  
FILING DATE: 14-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 812 amino acids  
TYPE: amino acid

STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-480-604A-29

Query Match 44.2%; Score 49.5; DB 1; Length 812;  
Best Local Similarity 52.2%; Pred. No. 2.6;  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

Oy 1 TGARTINGOLLYFRAN-GVQYKG 22  
Db 588 TGLRTIDGKKYFNTTAVATG 610

RESULT 12  
US-08-915-136-29  
Sequence 29, Application US/08915136  
Patent No. 6290960  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,136  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/480,604  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 812 amino acids  
TYPE: amino acid

```

; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-136-29

```

Query Match	44.28;	Score 49.5;	DB 4;	Length 812;
Best Local Similarity	52.28;	Pred. No. 2.6;		
Matches 12; Conservative	2;	Mismatches 8;	Indels 1;	Gaps 1;

```

Qy      1  TGARTINGQLLYFRAN-CVQVKC  22
          |||||: || |||
Db      588  TGLRTIDGKKYYFNTNTAVAVTG  610

```

RESULT 13  
US-08-480-C04A-6  
Sequence 6, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.  
 APPLICANT: THALLEY, BRUCE S.  
 APPLICANT: PADHAYE, NISHA V.  
 APPLICANT: FIRCA, JOSEPH R.  
 APPLICANT: STAFFORD, DOUGLAS C.  
 TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE  
 TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
 NUMBER OF SEQUENCES: 32  
 CORRESPONDENCE ADDRESS:

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/422,711  
FILING DATE: 14-APR-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763

INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2710 amino acids  
 TYPE: amino acid

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-480-604A-6

```

Query Match	44.2%;	Score 49.5;	DB 1;	Length 2710;
Best Local Similarity	52.2%;	Pred. No. 11;		
Matches 12;	Conservative 2;	Mismatches 8;	Indels 1;	Gaps 1;

```
QY      1  TGARTINGQLLYFRAN-GVQVKG  22
        |||||: || | |||
Db      2460 TGLRTIDGKKYYFNTNTAVAVTG  2482
```

RESULT 14  
US-08-405-496A-6  
; Sequence 6, Application US/08405496A  
; Patent No. 5919665  
; GENERAL INFORMATION:

APPLICANT: WILLIAMS, JAMES A.  
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
TITLE OF INVENTION: NEUROTOXIN  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 10 OCT 1994

FILING DATE: 02-DEC-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/985,321  
 FILING DATE: 04-DEC-1993

FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INCOLIA, DIANE E.  
REGISTRATION NUMBER: 40 027

TELEPHONE: (415) 397-8338  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:

MOLECULE TYPE: protein  
; US-08-405-496A-6

Query Match	44.2%	Score 49.5	DB 2	length 2710
Best Local Similarity	52.2%	Pred. No. 11		
Matches	12	Conservative	2	Mismatches 8; Indels 1; Gaps 1.
QY	1	TGARTINGQLLFRAN-GVQYKG	22	
		:		

Db 2460 TGLRTIDGKYYFNTNTAVATG 2482

## RESULT 15

US-08-915-136-6

Sequence 6, Application US/08915136  
Patent No. 6290960

## GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.

APPLICANT: THALLEY, BRUCE S.

APPLICANT: PADHIE, NISHA V.

APPLICANT: FIRCA, JOSEPH R.

APPLICANT: STAFORD, DOUGLAS C.

TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND

PREVENTION OF C. DIFFICILE DISEASE

NUMBER OF SEQUENCES: 32

## CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/915,136

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/480,604

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,496

FILING DATE: 16-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPD-01763

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-915-136-6

Query Match 44.2%; Score 49.5; DB 4; Length 2710;

Best Local Similarity 52.2%; Pred. No. 11;

Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

OY 1 TGARTINGOLLYFRAN-GVQVKG 22

Db 2460 TGLRTIDGKYYFNTNTAVATG 2482

Search completed: March 27, 2002, 13:59:28  
Job time: 561 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:23 ; Search time 1139.61 Seconds  
(without alignments)  
5.360 Million cell updates/sec

Title: US-09-290-049a-3  
Perfect score: 112  
Sequence: 1 TCGARTINGCOLLYFRANGVQVKG 22

Scoring table: BLOSOM62  
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main: \*  
1: /cgn2\_6/ptodata/2/paa/PCRUS\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep: \*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep: \*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep: \*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep: \*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep: \*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep: \*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep: \*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep: \*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep: \*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep: \*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep: \*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep: \*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep: \*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep: \*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep: \*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep: \*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep: \*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep: \*  
24: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	112	100.0	22	16	US-09-290-049-3
2	112	100.0	1475	19	US-09-557-848-2
3	112	100.0	1475	21	US-09-740-274-2
4	87	77.7	22	1	PCT-US93-04094-3
5	87	77.7	22	3	US-07-877-295-2
6	87	77.7	22	4	US-08-057-162-3
7	87	77.7	22	4	US-08-057-162-3
8	87	77.7	22	16	US-09-288-965-3
9	78	69.6	1375	21	US-09-740-274-4

10	67	59.8	1430	20	US-09-649-885-2	Sequence 2, Appli
11	67	59.8	1430	21	US-09-740-274-6	Sequence 6, Appli
12	65	58.0	2057	18	US-09-499-203-2	Sequence 2, Appli
13	49.5	44.2	320	18	US-09-446-269-1	Sequence 1, Appli
14	49.5	44.2	457	18	US-09-446-269-3	Sequence 3, Appli
15	49.5	44.2	811	1	PCT-US97-15394-7	Sequence 7, Appli
16	49.5	44.2	811	8	US-08-405-496-7	Sequence 7, Appli
17	49.5	44.2	811	8	US-08-422-711-7	Sequence 7, Appli
18	49.5	44.2	811	8	US-08-480-604-7	Sequence 7, Appli
19	49.5	44.2	811	13	US-08-704-159-7	Sequence 7, Appli
20	49.5	44.2	811	11	US-08-957-310-7	Sequence 7, Appli
21	49.5	44.2	811	14	US-09-084-517-7	Sequence 7, Appli
22	49.5	44.2	811	19	US-09-587-198-7	Sequence 7, Appli
23	49.5	44.2	812	8	US-08-422-711-29	Sequence 29, Appli
24	49.5	44.2	812	8	US-08-480-604-29	Sequence 29, Appli
25	49.5	44.2	812	14	US-09-084-517-29	Sequence 29, Appli
26	49.5	44.2	862	18	US-09-446-269-4	Sequence 4, Appli
27	49.5	44.2	866	19	US-09-545-772-2	Sequence 2, Appli
28	49.5	44.2	866	19	US-09-545-773-2	Sequence 2, Appli
29	49.5	44.2	2710	1	PCT-US97-15394-6	Sequence 6, Appli
30	49.5	44.2	2710	8	US-08-405-496-6	Sequence 6, Appli
31	49.5	44.2	2710	8	US-08-422-711-6	Sequence 6, Appli
32	49.5	44.2	2710	8	US-08-480-604-6	Sequence 6, Appli
33	49.5	44.2	2710	11	US-08-704-159-6	Sequence 6, Appli
34	49.5	44.2	2710	13	US-08-957-310-6	Sequence 6, Appli
35	49.5	44.2	2710	14	US-09-084-517-6	Sequence 6, Appli
36	49.5	44.2	2710	19	US-09-587-198-6	Sequence 6, Appli
37	47	42.0	785	18	US-09-488-125A-2538	Sequence 2538, Ap
38	47	42.0	785	19	US-09-538-092-872	Sequence 872, App
39	47	42.0	794	18	US-09-488-125A-6110	Sequence 6110, Ap
40	46	41.1	619	19	US-09-502-527-4	Sequence 4, Appli
41	46	41.1	619	22	US-09-882-774-1	Sequence 1, Appli
42	46	41.1	641	19	US-09-536-784-160	Sequence 160, App
43	46	41.1	641	21	US-09-765-271-160	Sequence 160, App
44	46	41.1	641	21	US-09-765-272-160	Sequence 160, App
45	46	41.1	648	6	US-08-247-491-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-290-049-3  
Sequence 3, Application US-09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: FDC98-012A  
CURRENT APPLICATION NUMBER: US-09/290,049  
CURRENT FILING DATE: 1999-04-12  
EARLIER APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115,142  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 22  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: GJB peptide  
US-09-290-049-3

Query Match 100.0%; Score 112; DB 16; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.7e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TCGARTINGCOLLYFRANGVQVKG 22  
|||||

Db 1 TGARTINGOLLYFRANGVQVGK 22

RESULT 2  
US-09-557-848-2

; Sequence 2, Application US/09557848  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0356D2  
; CURRENT APPLICATION NUMBER: US/09/557,848  
; CURRENT FILING DATE: 2000-04-26  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-557-848-2

Query Match 100.0%; Score 112; DB 19; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 2,1e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGVQVGK 22  
|||||  
Db 1300 TGARTINGOLLYFRANGVQVGK 1321

RESULT 3  
US-09-740-274-2

; Sequence 2, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 100.0%; Score 112; DB 21; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 2,1e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGVQVGK 22  
|||||

Db 1300 TGARTINGOLLYFRANGVQVGK 1321

RESULT 4  
PCT-US93-04094-3

; Sequence 3, Application PC/TUS9304094  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth Dental Infirmary for Children  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR  
; TITLE OF INVENTION: DENTAL  
; TITLE OF INVENTION: CARIES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Millitia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/04094  
; FILING DATE: 19930430  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia.  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: FDC92-01A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 22 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: Single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
PCT-US93-04094-3

Query Match 77.7%; Score 87; DB 1; Length 22;  
Best Local Similarity 77.3%; Pred. No. 3,1e-07;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGVQVGK 22  
|||||  
Db 1 TGAQTIKGOKLYFRANGVQVGK 22

RESULT 5  
US-07-877-295-2

; Sequence 2, Application US/07877295  
; GENERAL INFORMATION:  
; APPLICANT: Taudman, Martin A.  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Millitia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/877,295  
FILING DATE: 19920501  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Wagner, Richard W.  
REGISTRATION NUMBER: 34,480  
REFERENCE/DOCKET NUMBER: FDC92-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-877-295-2

Query Match 77.7%; Score 87; DB 3; Length 22;  
Best Local Similarity 77.3%; Pred. No. 3,1e-07;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGARTINGOLLYFRANGOVKG 22  
|||:| | | | | | | | | |  
Db 1 TGAQTIKGOKLYFRANGOVKG 22

RESULT 6  
US-08-057-162-3  
Sequence 3, Application US/08057162  
GENERAL INFORMATION:  
APPLICANT: Taudman, Martin A.  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR  
DENTAL  
TITLE OF INVENTION: CARRIES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/057,162  
FILING DATE: 19930430  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/877,295  
FILING DATE: 01-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wagner, Richard W.  
REGISTRATION NUMBER: 34,480  
REFERENCE/DOCKET NUMBER: FDC92-01A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-057-162A-3

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-057-162-3

Query Match 77.7%; Score 87; DB 4; Length 22;  
Best Local Similarity 77.3%; Pred. No. 3,1e-07;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGARTINGOLLYFRANGOVKG 22  
|||:| | | | | | | | | |  
Db 1 TGAQTIKGOKLYFRANGOVKG 22

RESULT 7  
US-08-057-162A-3  
Sequence 3, Application US/08057162A  
GENERAL INFORMATION:  
APPLICANT: Taudman, Martin A.  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR DENTAL  
DENTAL  
TITLE OF INVENTION: CARRIES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/057,162A  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/877,295  
FILING DATE: 01-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wagner, Richard W.  
REGISTRATION NUMBER: 34,480  
REFERENCE/DOCKET NUMBER: FDC92-01A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-057-162A-3

Query Match 77.7%; Score 87; DB 4; Length 22;  
Best Local Similarity 77.3%; Pred. No. 3,1e-07;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGARTINGOLLYFRANGOVKG 22  
|||:| | | | | | | | | |  
Db 1 TGAQTIKGOKLYFRANGOVKG 22

RESULT 8  
US-09-288-965-3  
Sequence 3, Application US/09288965  
GENERAL INFORMATION:  
APPLICANT: Lees, Andrew

```

RESULT 12
US-09-499-203-2
; Sequence 2, Application US/09499203
; GENERAL INFORMATION:

```

APPLICANT: KOSSMANN, Jens  
APPLICANT: WELSH, Thomas  
APPLICANT: QUANZ, Martin  
APPLICANT: KNUTH, Karola  
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
FILE REFERENCE: 147-196P  
CURRENT APPLICATION NUMBER: US/09/499,203  
CURRENT FILING DATE: 2000-02-08  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 2057  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-499-203-2

Query Match 58.0%; Score 65; DB 18; Length 2057;  
Best Local Similarity 57.1%; Pred. No. 0.3;  
Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

2 TGARTINGOLLYFRANGVOYKG 22  
|||:| | | | | | | | | |  
261 GLQRTIDGNLYFNOGVOYIKG 281

RESULT 13  
US-09-446-269-1  
Sequence 1, Application US/09446269  
GENERAL INFORMATION:  
APPLICANT: WARD, STEPHEN J.  
APPLICANT: WREN, BRENDAN W.  
APPLICANT: DOUGAN, GORDON  
APPLICANT: DOUCE, GILL  
TITLE OF INVENTION: IMMUNOGENIC FRAGMENTS OF TOXIN A OF CLOSTRIDIUM  
FILE REFERENCE: 550-166  
CURRENT APPLICATION NUMBER: US/09/446,269  
CURRENT FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/GB98/01805  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: GB 9713146.0  
PRIOR FILING DATE: 1997-06-20  
PRIOR APPLICATION NUMBER: GB 9800321.3  
PRIOR FILING DATE: 1998-01-07  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 320  
TYPE: PRT  
ORGANISM: Clostridium difficile  
9-446-269-1

Query Match 44.2%; Score 49.5; DB 18; Length 320;  
Best Local Similarity 52.2%; Pred. No. 16;  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

OY 1 TGARTINGOLLYFRAN-GVOYKG 22  
|||:| | | | | | | | | |  
Db 74 TGLRTIDGKRYFNTAVAVTG 96

RESULT 14  
US-09-446-269-3  
Sequence 3, Application US/09446269  
GENERAL INFORMATION:  
APPLICANT: WARD, STEPHEN J.  
APPLICANT: WREN, BRENDAN W.  
APPLICANT: DOUGAN, GORDON  
APPLICANT: DOUCE, GILL  
TITLE OF INVENTION: IMMUNOGENIC FRAGMENTS OF TOXIN A OF CLOSTRIDIUM  
FILE REFERENCE: 550-166  
CURRENT APPLICATION NUMBER: US/09/446,269  
CURRENT FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/GB98/01805  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: GB 9713146.0  
PRIOR FILING DATE: 1997-06-20  
PRIOR APPLICATION NUMBER: GB 9800321.3  
PRIOR FILING DATE: 1998-01-07  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 320  
TYPE: PRT  
ORGANISM: Clostridium difficile  
9-446-269-1

FILE REFERENCE: 550-166  
CURRENT APPLICATION NUMBER: US/09/446,269  
CURRENT FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/GB98/01805  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: GB 9713146.0  
PRIOR FILING DATE: 1997-06-20  
PRIOR APPLICATION NUMBER: GB 9800321.3  
PRIOR FILING DATE: 1998-01-07  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 457  
TYPE: PRT  
ORGANISM: Clostridium difficile  
US-09-446-269-3

Query Match 44.2%; Score 49.5; DB 18; Length 457;  
Best Local Similarity 52.2%; Pred. No. 23;  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

OY 1 TGARTINGOLLYFRAN-GVOYKG 22  
|||:| | | | | | | | | |  
Db 211 TGLRTIDGKRYFNTAVAVTG 233

RESULT 15  
PCT-US97-15394-7  
Sequence 7, Application PC/TUS9715394  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
APPLICANT: Thalley, Bruce S.  
TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
NUMBER OF SEQUENCES: 82  
CORRESPONDENCE ADDRESS:  
ADDRESS: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/15394  
FILING TYPE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingclia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-02304  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 811 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
PCT-US97-15394-7

Query Match 44.2%; Score 49.5; DB 18; Length 811;  
Best Local Similarity 52.2%; Pred. No. 45;  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

Wed Mar 27 15:11:54 2002

us-09-290-049a-3.rapm

Page 6

```
QY      1 TGARTINGQLLYFRAN-GVQVKG 22
          || |||: || | ||
Db      587 TGLRTIDGKKYYFNTNTAVAVTG 609
```

Search completed: March 27, 2002, 14:20:24  
Job time: 1571 sec

١٠٠

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:46 ; Search time 137.48 seconds  
(without alignments)  
11.042 Million cell updates/sec

Title: US-09-290-049a-3  
Perfect score: 112  
Sequence: 1 TGARTINGCOLLYFRANGVQVKG 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 68999538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	100.0	22	6	US-09-562-328-37 Sequence 37, Appl
2	112	100.0	22	6	US-09-290-049a-3 Sequence 3, Appl
3	87	77.7	22	6	US-09-562-328-7 Sequence 7, Appl
4	87	77.7	26	6	US-09-562-328-43 Sequence 43, Appl
5	49.5	44.2	811	7	US-10-011-366-7 Sequence 7, Appl
6	49.5	44.2	2710	7	US-10-011-366-6 Sequence 6, Appl
7	45	40.2	563	6	US-09-708-427-10858 Sequence 10858, A
8	45	40.2	630	6	US-09-708-427-10857 Sequence 10857, A
9	44	39.3	115	7	US-10-091-504-970 Sequence 970, App
10	44	39.3	148	5	US-09-621-976-4566 Sequence 4566, App
11	44	39.3	1675	6	US-09-708-427-30688 Sequence 30688, A
12	44	39.3	1684	6	US-09-708-427-30687 Sequence 30687, A
13	44	39.3	1780	6	US-09-708-427-30686 Sequence 30686, A
14	43.5	38.8	152	6	US-09-605-703B-1602 Sequence 1602, Ap
15	43	38.4	706	6	US-09-848-909-26 Sequence 26, Appl
16	43	38.4	875	6	US-09-848-909-32 Sequence 32, Appl
17	42	37.5	944	6	US-09-897-516-5617 Sequence 5617, Ap
18	42	37.5	929	5	US-09-748-875-60 Sequence 60, Appl
19	41.5	37.1	472	7	US-10-029-386-33685 Sequence 33685, A
20	41	36.6	174	6	US-09-675-784A-7970 Sequence 7970, Ap
21	41	36.6	242	6	US-09-897-516-4915 Sequence 4915, Ap
22	41	36.6	383	7	US-10-015-127-9902 Sequence 9902, App
23	41	36.6	474	6	US-09-604-693A-106 Sequence 106, App
24	41	36.6	846	8	US-60-341-261-3302 Sequence 3302, Ap
25	40.5	36.2	428	7	US-10-015-127-13123 Sequence 13123, A

26	40	35.7	19	6	US-09-562-328-39 Sequence 39, Appl
27	40	35.7	169	6	US-09-605-703B-1160 Sequence 1160, Ap
28	40	35.7	231	6	US-09-631-616-17 Sequence 17, Appl
29	40	35.7	266	6	US-09-708-427-17269 Sequence 17269, A
30	40	35.7	363	6	US-09-708-427-17268 Sequence 17268, A
31	40	35.7	379	6	US-09-614-150-33408 Sequence 33408, A
32	40	35.7	400	6	US-09-605-703B-1178 Sequence 1178, Ap
33	40	35.7	441	6	US-09-708-427-17267 Sequence 17267, A
34	40	35.7	590	6	US-09-708-427-9652 Sequence 9652, Ap
35	40	35.7	638	6	US-09-708-427-9651 Sequence 9651, Ap
36	40	35.7	669	6	US-09-708-427-9651 Sequence 9651, Ap
37	39.5	35.3	304	7	US-10-015-127-11175 Sequence 11175, A
38	39	34.8	91	7	US-10-011-366-8 Sequence 8, Appl
39	39	34.8	137	6	US-09-614-150-15900 Sequence 15900, A
40	39	34.8	196	6	US-09-708-427-9649 Sequence 9649, Ap
41	39	34.8	219	6	US-09-708-427-9648 Sequence 9648, Ap
42	39	34.8	319	6	US-09-708-427-9645 Sequence 9645, Ap
43	39	34.8	351	6	US-09-675-784A-8859 Sequence 8859, Ap
44	39	34.8	371	6	US-09-708-427-9647 Sequence 9647, Ap
45	39	34.8	393	7	US-10-015-127-12338 Sequence 12338, A

## ALIGNMENTS

```
RESULT 1.
US-09-562-328-37
: Sequence 37, Application US/09562328
: GENERAL INFORMATION:
: APPLICANT: LEEB, ANDREW
: APPLICANT: TAUBMAN, MARTIN A.
: TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
: FILE REFERENCE: 04995.0046-01
: CURRENT FILING DATE: US/09/562.328
: PRIOR APPLICATION NUMBER: 2000-05-01
: PRIOR FILING DATE: 1999-04-09
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: Patent Ver. 2.1
: SEQ ID NO 37
: LENGTH: 22
: TYPE: PRT
: ORGANISM: Streptococcus sp.
US-09-562-328-37

Query Match      100.0%; Score 112; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.8e-12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGARTINGCOLLYFRANGVQVKG 22
|||||
Db      1 TGARTINGCOLLYFRANGVQVKG 22

RESULT 2
US-09-290-049a-3
: Sequence 3, Application US/09290049A
: GENERAL INFORMATION:
: APPLICANT: Smith, Daniel J.
: APPLICANT: Taubman, Martin A.
: TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
: FILE REFERENCE: 1564.1008-002
: CURRENT FILING DATE: US/09/290.049A
: PRIOR APPLICATION NUMBER: 1999-04-12
: PRIOR FILING DATE: 1998-04-13
: PRIOR APPLICATION NUMBER: 60/081.550
: PRIOR FILING DATE: 1998-04-13
: PRIOR APPLICATION NUMBER: 60/115.142
: PRIOR FILING DATE: 1999-01-08
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 4.0
```

Query Match	77.7%;	Score 87;	DB 6;	Length 26;
*Best Local Similarity	77.3%;	Pred. No.	5.3e-08;	

RESULT 6  
US-10-011-366-6

; Sequence 6, Application US/10011366  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; Kink, John A.  
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
; DISEASE  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/011,366  
; FILING DATE: 16-Nov-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/957,310  
; FILING DATE: 23-OCT-1997  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 24-OCT-1994  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPND-01121  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2710 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-011-366-6  
Query Match 44.2%; Score 49.5; DB 7; Length 2710;  
Best Local Similarity 52.2%; Pred. No. 25;  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;  
OY 1 TGARTINGOLLYFRAN-GVQYKG 22  
DB 2460 TGLRTIDGKRYFTNTAVATG 2482  
RESULT 7  
US-09-708-427-10858  
; Sequence 10858, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. Alexandrov et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708,427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 10858  
; LENGTH: 563  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: 1..563  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc-feature  
; LOCATION: 1..563  
; OTHER INFORMATION: Ceres Seq. ID 1822113  
US-09-708-427-10858  
Query Match 40.2%; Score 45; DB 6; Length 563;  
Best Local Similarity 45.5%; Pred. No. 22;  
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
OY 1 TGARTINGOLLYFRANGVQYKG 22  
DB 72 TGIGPIHGRICFMFMANDPTVKG 93  
RESULT 8  
US-09-708-427-10857  
; Sequence 10857, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. Alexandrov et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708,427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10857  
; LENGTH: 630  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: 1..630  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc-feature  
; LOCATION: 1..630  
; OTHER INFORMATION: Ceres Seq. ID 1822112  
US-09-708-427-10857  
Query Match 40.2%; Score 45; DB 6; Length 630;  
Best Local Similarity 45.5%; Pred. No. 26;  
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
OY 1 TGARTINGOLLYFRANGVQYKG 22  
DB 139 TGIGPIHGRICFMFMANDPTVKG 160  
RESULT 9  
US-10-091-504-970  
; Sequence 970, Application US/10091504  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC007C1  
; CURRENT APPLICATION NUMBER: US/10/091,504  
; CURRENT FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 2442  
; Prior Application removed - See File Wrapper or Palm  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 970  
; LENGTH: 115  
; TYPE: PRT

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (26)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (28)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (30)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (46)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (50)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (59)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-091-504-970
```

```

Query Match          39.3%; Score 44; DB 7; Length 115;
Best Local Similarity 70.0%; Pred. No. 5.1;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```

5 TINGOLLYR 14
1:11:11:11
63 TVNGELLFR 72
```

```

RESULT 10
US-09-621-976-4566
; Sequence 4566, Application US/09621976
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 4566
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-621-976-4566
```

```

Query Match          39.3%; Score 44; DB 5; Length 148;
Best Local Similarity 70.0%; Pred. No. 6.8;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```

QY          5 TINGOLLYR 14
1:11:11:11
DB          63 TVNGELLFR 72
```

```

RESULT 11
US-09-708-427-30688
; Sequence 30688, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30688
; LENGTH: 1675
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1675
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..1675
; OTHER INFORMATION: Ceres Seq. ID 1830036
; US-09-708-427-30688
```

```

Query Match          39.3%; Score 44; DB 6; Length 1675;
Best Local Similarity 53.3%; Pred. No. 1.2e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
```

```

QY          1 TCGARTINGOLLYERA 15
1:11:1:1:1:11
DB          881 TLAIRYRGMMYYRA 895
```

```

RESULT 12
US-09-708-427-30687
; Sequence 30687, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30687
; LENGTH: 1684
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1684
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..1684
; OTHER INFORMATION: Ceres Seq. ID 1830035
; US-09-708-427-30687
```

```

Query Match          39.3%; Score 44; DB 6; Length 1684;
Best Local Similarity 53.3%; Pred. No. 1.2e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY          1 TCGARTINGOLLYERA 15
1:11:1:1:1:11
```



Db 890 TLARTVRCGMYYRA 904

## RESULT 13

US-09-708-427-30686  
; Sequence 30686, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708.427  
; PRIOR FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 30686  
; LENGTH: 1780  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 1..1780  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc.feature  
; LOCATION: 1..1780  
; OTHER INFORMATION: Ceres Seq. ID 1830034  
US-09-708-427-30686

## Query Match

Best Local Similarity 39.3%; Score 44; DB 6; Length 1780;  
Best Local Similarity 53.3%; Pred. No. 1.3e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGARTINGQLYFRA 15

Db 986 TLARTVRCGMYYRA 1000

## RESULT 14

US-09-605-703B-1602  
; Sequence 1602, Application US/09605703B  
; GENERAL INFORMATION:  
; APPLICANT: Pompeius, Markus  
; APPLICANT: Krieger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauser, Gregor  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL  
; FILE REFERENCE: BGI-129CP  
; CURRENT APPLICATION NUMBER: US/09/605.703B  
; PRIOR FILING DATE: 2000-06-27  
; PRIOR APPLICATION NUMBER: 60/142,764  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: 60/152,318  
; PRIOR FILING DATE: 1999-09-03  
; NUMBER OF SEQ ID NOS: 2934  
; SEQ ID NO 1602  
; LENGTH: 152  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-605-703B-1602

## Query Match

Best Local Similarity 38.8%; Score 43.5; DB 6; Length 152;  
Best Local Similarity 54.5%; Pred. No. 8.5;  
Matches 12; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 1 TGARTINGQLYFRANGVQYK 22

Db 125 TGA-TISASANYLRAAGVQVRG 145

## RESULT 15

US-09-848-909-26  
; Sequence 26, Application US/09848909  
; GENERAL INFORMATION:  
; APPLICANT: Collier, R. John  
; APPLICANT: Sellman, Brett R.  
; TITLE OF INVENTION: Compounds and Methods for the Treatment  
; FILE REFERENCE: 00742/060002  
; CURRENT APPLICATION NUMBER: US/09/848.909  
; PRIOR FILING DATE: 2001-05-04  
; PRIOR APPLICATION NUMBER: US 60/201,800  
; PRIOR FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PastSeq for Windows Version 4.0  
; SEQ ID NO 26  
; LENGTH: 706  
; TYPE: PRT  
; ORGANISM: Clostridium perfringens  
US-09-848-909-26

## Query Match

Best Local Similarity 38.4%; Score 43; DB 6; Length 706;  
Best Local Similarity 56.2%; Pred. No. 63;  
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 5 TINGQLYFRANGVQY 20

Db 388 TKNGLYF--NGIRP 401

Search completed: March 27, 2002, 14:22:46  
Job time: 1693 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:16 ; Search time 102.51 Seconds  
(without alignments)  
16.348 Million cell updates/sec

Title: US-09-290-049a-3  
Perfect score: 112  
Sequence: 1 TCARTINGQLTYFRANGVQVKG 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	100.0	1475	2 B33135	glfB protein precu
2	87	77.7	1592	2 A38175	glucosyltransferas
3	82	73.2	1599	2 S22737	glucosyltransferas
4	80	71.4	1518	2 A44811	glucosyltransferas
5	78	69.6	1375	2 J70345	dextranucrase (EC
6	75	67.0	1290	2 J05473	dextranucrase (EC
7	75	67.0	1365	2 A41483	glucosyltransferas
8	73	65.2	1449	2 T30857	glucosyltransferas
9	72	64.3	1508	2 T31098	probable dextranu
10	68	60.7	1449	2 T30552	glucan-binding pro
11	68	59.8	1431	2 A45866	glucosyltransferas
12	67	59.8	1577	2 T30858	dextranucrase (EC
13	66	58.9	1577	2 T30858	dextranucrase (EC
14	65.5	58.5	329	2 A55221	dextranase inhibito
15	54	48.2	221	2 T37244	GATA transcription
16	50	44.6	831	2 T00323	chitinase (EC 3.2.
17	49.5	44.2	529	2 T10388	hypothetical prote
18	49.5	44.2	2710	2 A37052	toxin A - Clostrid
19	47	42.0	126	2 S40016	doc protein - phag
20	47	42.0	696	2 A29635	transcriptin fact
21	47	42.0	788	2 J50747	regulatory protein
22	46	41.1	231	2 F70471	conserved hypothet
23	46	41.1	245	2 F64465	hypothetical prote
24	46	41.1	301	2 S13165	astaloglycoprotein
25	46	41.1	619	2 A41971	surface protein ps
26	46	41.1	1301	2 S18118	alpha-amyrase - Al
27	46	41.1	2364	2 I40884	cytotoxin L - Clo
28	45	40.2	630	2 T05433	hypothetical prote
29	45	40.2	648	2 S10869	enterotoxin A - Cl

30	44	39.3	329	2 F71014	probable dipeptide
31	44	39.3	522	2 D71074	probable methylal
32	44	39.3	522	2 H75135	methylmalonyl-coa
33	44	39.3	566	2 S74633	high affinity sulf
34	44	39.3	591	2 C84220	propionyl-CoA carb
35	44	39.3	829	2 T29372	hypothetical prote
36	44	39.3	1396	2 S36851	L-shaped tail fibe
37	44	39.3	1679	2 T30271	surface protein -
38	44	39.3	1780	2 A85045	probable glucan sy
39	44	39.3	2178	2 S5805	alpha-toxin - Clo
40	43	38.4	110	2 B69403	hypothetical prote
41	43	38.4	299	2 T15928	hypothetical prote
42	43	38.4	305	2 B75474	probable chloride
43	43	38.4	329	2 H75117	peptide abc transp
44	43	38.4	467	1 S45493	serine proteinase
45	43	38.4	509	2 A49094	methylmalonyl-CoA

## ALIGNMENTS

```

RESULT 1
B33135
glfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C>Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 15-Oct-1999
C:Accession: B33135, A33128
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the glfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013
A:Accession: B33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SH2>
A:Cross-references: GB:M17361; NID:G153639; PIDN:AAA8588.1; PID:G153640
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
submitted to the Protein Sequence Database, September 1990
A:Reference number: A33128
A:Accession: A33128
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-171, 173-641, 'N', 643-1475 <SH2>
A:Experimental source: strain GS-5
C:Superfamily: cpl repeat homology
F:1096-1115/Domain: cpl repeat homology <CP1>
F:1224-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>
F:1354-1373/Domain: cpl repeat homology <CP4>
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 100.0%; Score 112; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 4.9e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCARTINGQLTYFRANGVQVKG 22
Db 1300 TCARTINGQLTYFRANGVQVKG 1321

RESULT 2
A38175
glucosyltransferase precursor - Streptococcus sobrius
C:Species: Streptococcus sobrius
C>Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Oct-1999
C:Accession: A38175
R:Abu, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
J. Bacteriol. 173, 989-996, 1991
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococ
A:Reference number: A38175; MUID:91123227
A:Accession: A38175
A>Status: preliminary

```

A:Molecule type: DNA  
 A:Residues: 1-1592 <ABO>  
 A:Cross-references: GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID:01014946; PID:g217033  
 A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBI:81052)  
 C:Genetics:  
 A:Gene: gtfJ  
 C:Superfamily: cpl repeat homology  
 C:Keywords: glycosyltransferase; hexosyltransferase  
 F.1307-1326/Domain: cpl repeat homology <CP4>

Query Match 77.7%; Score 87; DB 2; Length 1592;  
 Best Local Similarity 77.3%; Pred. No. 7e-06;  
 Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

1 TGARTINGOLLYFRANGOVKG 22  
 |||:|||||:|||||:|||||  
 Db 1298 TGAQRTIGCKLYFRANGOVKG 1319

RESULT 3  
 S22737  
 C:Species: Streptococcus salivarius  
 C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
 C:Accession: S22737; S28810; B44811; S22727  
 R:Jacques, N.  
 A:Reference number: S22726  
 A:Accession: S22737  
 A:Molecule type: DNA  
 A:Residues: 1-1599 <JAC>  
 A:Cross-references: EMBL:211872; NID:g47530; PIDN:CAA77898.1; PID:g47531  
 A:Experimental source: ATCC 25975  
 R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
 J. Gen. Microbiol. 137, 2577-2593, 1991  
 A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
 A:Reference number: A44811; MUID:92148377  
 A:Accession: S28810  
 A:Molecule type: DNA  
 A:Residues: 1-51 <GIF>  
 A:Cross-references: EMBL:211873  
 A:Genetics:  
 A:Gene: gtfK  
 C:Superfamily: cpl repeat homology  
 C:Keywords: glycosyltransferase; hexosyltransferase  
 F.1456-1475/Domain: cpl repeat homology <CPR>

Query Match 73.2%; Score 82; DB 2; Length 1599;  
 Best Local Similarity 77.3%; Pred. No. 4.7e-05;  
 Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

1 TGARTINGOLLYFRANGOVKG 22  
 |||:|||||:|||||:|||||  
 Db 1402 TGAQVINGOLLYFRANGOVKG 1423

RESULT 4  
 A44811  
 C:Species: Streptococcus salivarius  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999  
 R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
 J. Gen. Microbiol. 137, 2577-2593, 1991  
 A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
 A:Reference number: A44811; MUID:92148377  
 A:Accession: A44811  
 A:Molecule type: DNA

A:Residues: 1-1518 <GIF>  
 A:Cross-references: EMBL:211873; NID:g47526; PIDN:CAA77900.1; PID:g47527  
 A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBI:81052)  
 C:Genetics:  
 A:Gene: gtfJ  
 C:Superfamily: cpl repeat homology  
 C:Keywords: glycosyltransferase; hexosyltransferase  
 F.1307-1326/Domain: cpl repeat homology <CP4>

Query Match 71.4%; Score 80; DB 2; Length 1518;  
 Best Local Similarity 72.7%; Pred. No. 9.5e-05;  
 Matches 16; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

1 TGARTINGOLLYFRANGOVKG 22  
 |||:|||||:|||||:|||||  
 Db 1383 TGAQVINGOLLYFRANGOVKG 1404

RESULT 5  
 J70345  
 C:Species: Streptococcus mutans  
 C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
 R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
 Gene 69, 101-109, 1988  
 A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.  
 A:Reference number: J70345; MUID:89137980  
 A:Accession: J70345  
 A:Molecule type: DNA  
 A:Residues: 1-1375 <UED>  
 A:Experimental source: GS-5  
 R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
 J. Bacteriol. 169, 4263-4270, 1987  
 A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
 A:Reference number: A33135; MUID:87308013  
 A:Accession: C33135  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-349 <SHI>  
 A:Cross-references: GB:M17361  
 C:Genetics:  
 A:Gene: gtfC  
 C:Function: catalyzes the synthesis of both water-soluble and water-insoluble gluc  
 C:Superfamily: cpl repeat homology  
 C:Keywords: duplication; glycosyltransferase; hexosyltransferase  
 F.1-34/Domain: signal sequence #status predicted <SIG>  
 F.35-1375/Product: glucosyltransferase #status predicted <MAT>  
 F.1126-1145/Domain: cpl repeat homology <CP1>  
 F.1125-1272/Domain: cpl repeat homology <CP2>  
 F.1318-1337/Domain: cpl repeat homology <CP3>

Query Match 69.6%; Score 78; DB 2; Length 1375;  
 Best Local Similarity 68.2%; Pred. No. 0.00018;  
 Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

1 TGARTINGOLLYFRANGOVKG 22  
 |||:|||||:|||||:|||||  
 Db 1264 TGTVTINCOLLYFRANGOVKG 1285

RESULT 6  
 J55473  
 C:Species: Leuconostoc mesenteroides  
 C:Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997  
 R:Monchois, V.; Willems, R.M.; Renaud-Slmeon, M.; Croux, C.; Monsan, P.  
 Gene 182, 23-32, 1996

A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from *Leuconostoc mesenteroides*  
A:Reference number: J05473; MUID:9713686  
A:Accession: J05473  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1290 <MON>  
A:Cross-references: GB:008181  
C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto  
C:Genetics:  
A:Gene: dsra  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:78-870/Domain: catalytic #status predicted <CAT>  
F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 67.0%; Score 75; DB 2; Length 1290;  
Best Local Similarity 68.2%; Pred. No. 0.00053;  
Matches 15; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGARTINGOLLYFRANGOVKG 22  
||:|||||||:|  
1245 TGVVINGOLLYFDADGRQVKG 1266

RESULT 7  
A1483  
glucosyltransferase (EC 2.4.1.-) gtfS precursor - *Streptococcus sobrinus*  
C:Species: *Streptococcus sobrinus*  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C:Accession: A1483  
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
A:Title: Analysis of the *Streptococcus downei* gtfS gene, which specifies a glucosyltrans  
A:Reference number: A1483; MUID:9031665  
A:Accession: A1483  
A:Molecule type: DNA  
A:Residues: 1-1365 <GIL>  
A:Cross-references: GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:g153653  
C:Genetics:  
A:Gene: gtfS  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 67.0%; Score 75; DB 2; Length 1365;  
Best Local Similarity 63.6%; Pred. No. 0.00057;  
Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGARTINGOLLYFRANGOVKG 22  
||:|||||||:|  
1235 TGEQTIDGKVFQDNGVQVKG 1256

RESULT 8  
T30857  
glucosyltransferase - *Streptococcus salivarius*  
C:Species: *Streptococcus salivarius*  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30857  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
A:Title: Immun. 63, 609-621, 1995  
A:Reference number: 220909; MUID:95122197  
A:Accession: T30857  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <SIM>  
A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AAC41412.1  
C:Genetics:  
A:Gene: gtfI

Query Match 65.2%; Score 73; DB 2; Length 1449;

Best Local Similarity 68.2%; Pred. No. 0.0013;  
Matches 15; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGARTINGOLLYFRANGOVKG 22  
||:|||||||:|  
DB 1258 TGHONINGELFFDNGVQVKG 1279

RESULT 9  
T31098  
probable dextranucrase (EC 2.4.1.5), extracellular - *Leuconostoc mesenteroides*  
C:Species: *Leuconostoc mesenteroides*  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C:Accession: T31098  
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase  
A:Reference number: 220981; MUID:98164374  
A:Accession: T31098  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1508 <MON>  
A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1  
A:Experimental source: strain NRRL B-1299  
C:Genetics:  
A:Gene: dsrb  
C:Function:  
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 64.3%; Score 72; DB 2; Length 1508;  
Best Local Similarity 68.2%; Pred. No. 0.002;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGARTINGOLLYFRANGOVKG 22  
||:|||||||:|  
DB 1399 TGLVINGNLKVFQANGRVKG 1420

RESULT 10  
A37184  
glucan-binding protein - *Streptococcus mutans*  
C:Species: *Streptococcus mutans*  
C:Date: 31-Jan-1992 #sequence\_revision 31-Jan-1992 #text\_change 15-Oct-1999  
C:Accession: A37184  
R:Banas, J.A.; Russell, R.R.B.; Ferretti, J.J.  
A:Title: Immun. 58, 667-673, 1990  
A:Reference number: A37184; MUID:90170123  
A:Accession: A37184  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-561 <BAN>  
A:Cross-references: GB:M30945; NID:g153637; PIDN:AAA26894.1; PID:g153638  
C:Superfamily: cpl repeat homology  
F:169-188/Domain: cpl repeat homology <CP1>  
F:164-283/Domain: cpl repeat homology <CP2>  
F:349-368/Domain: cpl repeat homology <CP3>  
F:504-523/Domain: cpl repeat homology <CP4>  
F:525-548/Domain: cpl repeat homology <CP5>

Query Match 60.7%; Score 68; DB 2; Length 563;  
Best Local Similarity 66.7%; Pred. No. 0.0031;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 GARTINGOLLYFRANGOVKG 22  
| |||||:| |||||  
DB 276 GMRITGKRVYFPTNGVQVKG 296

RESULT 11

T30552 glucosyltransferase N - *Streptococcus salivarius* (fragment)  
C:Species: *Streptococcus salivarius*  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30552  
R:Jaffe, R.I.  
submitted to the EMBL Data Library, February 1998  
A:Description: *Streptococcus salivarius* V1477 gtfN.  
A:Reference number: Z20854  
A:Accession: T30552  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <JAF>  
A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1  
C:Genetics:  
A:Gene: gtfN

Query Match	Score	DB	Length
Best Local Similarity	66.7%	Pred. No. 0.0086	
Matches	14	Conservative	2
		Mismatches	5
		Indels	0
		Gaps	0

```

RESULT 12
A45866
dextran sucrose (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: A45866
R:Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl
A:Reference number: A45866; MUID:91100958
A:Accession: A45866
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: GB:M29296
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <Cpl>
F:127-1146/Domain: cpl repeat homology <CP2>
F:192-1211/Domain: cpl repeat homology <CP3>
F:11257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP8>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1365-1404/Domain: cpl repeat homology <CP7>

```

	Query Match	Best Local Similarity	Matches	Score	DB 2	Length	1431
				59.8%	59.1%		
					Pred. No.	0.012	
					5	Mismatches	4
						Indels	0
						Gaps	0
Qy	1	TGARTINGOLLYFRANGVYK	22				
		11::11111::11::11111					
Db	1332	TGSGTIAAGKLLFASDGGKVG	1353				

```

RESULT 13
T30858
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30858
R:Simpson, C.L.: Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 2575 possesses at least two genes coding for
A:Reference number: Z20909; MUID:95122197

```

A:Accession:J130858  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1577 <SIM>  
 A:Cross-references: EMBL:L35928, NID:g662380, PID:g662381, PIDN:AMC41413.1  
 C:Genetics:  
 A:Gene: gltm

Query Match	58.9%	Score 66:	DB 2:	Length 1577:
Best Local Similarity	54.5%	Pred. No. 0.02:		
Matches 12:	Conservative	4:	Mismatches	6:
			Indels	0:
			Gaps	0:
QY	1	TGARTINGQLYFRANGVQYKG	22	
Db	1469	TGQINNNKVVYFGSNGAQVKG	1490	

```

RESULT 14
A55221
dextranase inhibitor precursor - Streptococcus sobrinus (strain UAB65, serotype g)
C:Species: Streptococcus sobrinus
C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 17-Mar-1995
C:Accession: A55221
R:Sun, J.W.; Wanda, S.Y.; Camilli, A.; Curtiss III, R.
J. Bacteriol. 176, 7213-7222, 1994
A:Title: Cloning and DNA sequencing of the dextranase inhibitor gene (dei) from Strept
A:Reference number: A55221; MUID:95050304
A:Accession: A55221
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-329 <SUN>
A:Cross-references: GB:I34406
C:Superfamily: cpl repeat homology
F:175-192/Domain: cpl repeat homology <CP1>

```

Query Match	58.5%	Score 65.5	DB 2	Length 329
Best Local Similarity	68.2%	Pred. No. 0.0044		
Matches 15; Conservative	1;	Mismatches 5;	Indels 1;	Gaps 1
QY	2	GARTINGOLLYFRAN-GVOYKG	22	
		:		
DB	165	GLÖTINGÖLLFYDNINGKOVKG	206	

RESULT 15

T37244

CAT:transcription factor end-1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000

C:Accession: T37244; T22912

C:Zhu, J.; Hill, R.J.; Heid, P.J.; Fukuyama, M.; Sugimoto, A.; Pless, J.R.; Rothman

Genes Dev. 11, 2883-2896, 1997

A:Title: End-1 encodes an apparent GATA factor that specifies the endoderm precursor

A:Reference number: Z21642; MUID:96019168

A:Accession: T37244

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-221 <ZH1>

A:Cross-references: EMBL:AF026555; NID:g2655205; PIDN:AAB97516.1; PID:g2655206

R:McMurray, A.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19637

A:Accession: T22912

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-221 <W1>

A:Cross-references: EMBL:Z81555; PIDN:CA804513.1; GSPBD:GND0023; CESP:F58E10.2

A:Experimental source: clone F58E10

C:Genetics:

A:Gene: end-1; F58E10.2

A:Map position: 5

A: Introns: 71/1; 128/3; 162/2

C: Function:

A: Description: involved in the specification of germ layer and founder cell identity

C: Keywords: transcription factor; zinc finger

Query Match

48.2%; Score 54; DB 2; Length 221;

Best Local Similarity 57.9%; Pred. No. 0.23;

Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGARTINGQLYFRANGVQ 19

DB 156 SGAIECNCGLYFRKNGIQ 174

Search completed: March 27, 2002, 14:01:17  
JOD Time: 479 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:54:13 ; Search time 1139.61 seconds  
(without alignments)  
5.116 Million cell updates/sec

Title: US-09-290-049a-1

Sequence: 1 ANDHLSILEAWSNDPTPLD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/PCRTUS\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep:\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep:\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep:\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep:\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep:\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep:\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep:\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep:\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep:\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep:\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep:\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep:\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep:\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep:\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep:\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep:\*  
24: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	21	16	US-09-290-049-1
2	119	100.0	21	16	US-09-290-049-10
3	119	100.0	1475	19	US-09-557-848-2
4	119	100.0	1475	21	US-09-740-274-2
5	110	90.4	1375	21	US-09-740-274-4
6	108	90.8	21	16	US-09-290-049-12
7	108	90.8	21	16	US-09-290-049-14
8	75	63.0	21	16	US-09-290-049-11
9	75	63.0	1430	20	US-09-649-885-2

10	75	63.0	1430	21	US-09-740-274-6	Sequence 6, Appl1
11	73	61.3	21	16	US-09-290-049-13	Sequence 13, Appl1
12	60	50.4	2057	18	US-09-499-203-2	Sequence 2, Appl1
13	50	42.0	306	19	US-09-595-298A-661	Sequence 61, App
14	50	42.0	306	19	US-09-595-298A-660	Sequence 60, App
15	49.5	41.6	546	24	US-60-324-109-17015	Sequence 17015, A
16	49.5	41.6	549	24	US-60-324-109-16789	Sequence 16789, A
17	49.5	41.2	391	24	US-60-324-109-20736	Sequence 20736, A
18	48.5	40.8	401	1	PCT-US01-01309-222	Sequence 222, App
19	48	40.3	773	16	US-09-248-796-22573	Sequence 22573, A
20	47.5	39.9	178	22	US-09-867-716-19848	Sequence 19848, A
21	47.5	39.9	534	17	US-09-344-882-24	Sequence 24, Appl1
22	47.5	39.9	534	24	US-60-324-109-33149	Sequence 33149, A
23	47	39.5	196	19	US-09-595-298A-765	Sequence 765, App
24	46.5	39.1	220	24	US-60-212-413-164	Sequence 164, App
25	46.5	39.1	220	24	US-60-229-518-331	Sequence 331, App
26	46.5	39.1	330	24	US-60-207-422-75	Sequence 75, Appl1
27	46.5	39.1	539	1	PCT-US01-01332-762	Sequence 762, App
28	46.5	39.1	539	21	US-09-764-875-762	Sequence 762, App
29	46	38.7	62	18	US-09-417-507-26597	Sequence 26597, A
30	46	38.7	94	1	PCT-US01-08656-5595	Sequence 5595, Ap
31	46	38.7	118	20	US-09-673-958A-221	Sequence 221, App
32	46	38.7	199	21	US-09-738-626-4571	Sequence 4571, Ap
33	46	38.7	275	19	US-09-538-092-348	Sequence 348, App
34	46	38.7	336	15	US-09-198-452A-662	Sequence 662, App
35	46	38.7	336	18	US-09-438-185-626	Sequence 626, App
36	46	38.7	394	21	US-09-739-449-8822	Sequence 8822, Ap
37	46	38.7	394	22	US-09-803-110-8822	Sequence 8822, Ap
38	46	38.7	476	1	PCT-US01-22636-9	Sequence 9, Appl1
39	46	38.7	476	24	US-60-219-231-9	Sequence 9, Appl1
40	46	38.7	596	16	US-09-252-991A-21255	Sequence 21255, A
41	46	38.7	2590	24	US-60-215-161-7203	Sequence 7203, Ap
42	46	38.7	2603	24	US-60-215-161-5891	Sequence 5891, Ap
43	45	37.8	185	24	US-60-215-161-5719	Sequence 5719, Ap
44	45	37.8	194	1	PCT-US01-14827-9825	Sequence 9825, Ap
45	45	37.8	236	15	US-09-154-750-80	Sequence 80, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-290-049-1  
Sequence 1, Application US/09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: PDC98-0192A  
CURRENT APPLICATION NUMBER: US/09/290, 049  
EARLIER FILING DATE: 1999-04-12  
EARLIER APPLICATION NUMBER: 60/081, 550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115, 142  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: FAW peptide  
US-09-290-049-1

Query Match 100.0%; Score 119; DB 16; Length 21;  
Best Local Similarity 100.0%; Pred. No. 5.3e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 ANDHLSILEAWSNDPTPLD 21  
|||||

DB 1 ANDHLSILEAWSNDNTPYLHD 21

## RESULT 2

US-09-290-049-10  
; Sequence 10, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-0192A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049-10

Query Match 100.0%; Score 119; DB 16; Length 21;  
Best Local Similarity 100.0%; Pred. No. 5.3e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDNTPYLHD 21  
Db 1 ANDHLSILEAWSNDNTPYLHD 21

## RESULT 3

US-09-557-848-2  
; Sequence 2, Application US/09557848  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 0356D2  
; CURRENT APPLICATION NUMBER: US/09/557,848  
; PRIOR FILING DATE: 2000-04-26  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-557-848-2

Query Match 100.0%; Score 119; DB 19; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 5.2e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDNTPYLHD 21  
Db 481 ANDHLSILEAWSNDNTPYLHD 501

## RESULT 4

US-09-740-274-2  
; Sequence 2, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Glucan-containing Compositions and Paper

; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 100.0%; Score 119; DB 21; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 5.2e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDNTPYLHD 21  
Db 481 ANDHLSILEAWSNDNTPYLHD 501

## RESULT 5

US-09-740-274-4  
; Sequence 4, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 92.4%; Score 110; DB 21; Length 1375;  
Best Local Similarity 95.2%; Pred. No. 1.2e-07;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDNTPYLHD 21

Db 507 ANDHLSILEAMSDNDPTYLHD 527

RESULT 6

US-09-290-049-12

; Sequence 12, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT APPLICATION NUMBER: US/09/290,049

; EARLIER FILING DATE: 1999-04-12

; EARLIER APPLICATION NUMBER: 60/081,550

; EARLIER FILING DATE: 1998-04-13

; EARLIER APPLICATION NUMBER: 60/115,142

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 12

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. downei

US-09-290-049-12

Query Match 90.8%; Score 108; DB 16; Length 21;

Best Local Similarity 85.7%; Pred. No. 2.7e-09;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDPTYLHD 21

Db 1 ANNHVSIVEAMSDNDPTYLHD 21

RESULT 7

US-09-290-049-14

; Sequence 14, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT APPLICATION NUMBER: US/09/290,049

; EARLIER FILING DATE: 1999-04-12

; EARLIER APPLICATION NUMBER: 60/081,550

; EARLIER FILING DATE: 1998-04-13

; EARLIER APPLICATION NUMBER: 60/115,142

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 14

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. sobrinus

US-09-290-049-14

Query Match 90.8%; Score 108; DB 16; Length 21;

Best Local Similarity 85.7%; Pred. No. 2.7e-09;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDPTYLHD 21

Db 1 ANNHVSIVEAMSDNDPTYLHD 21

RESULT 8

US-09-290-049-11

; Sequence 11, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT APPLICATION NUMBER: US/09/290,049

; EARLIER FILING DATE: 1999-04-12

; EARLIER APPLICATION NUMBER: 60/081,550

; EARLIER FILING DATE: 1998-04-13

; EARLIER APPLICATION NUMBER: 60/115,142

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 11

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. mutans

US-09-290-049-11

Query Match 63.0%; Score 75; DB 16; Length 21;

Best Local Similarity 71.4%; Pred. No. 0.00036;

Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDPTYLHD 21

Db 1 AINHLSILEAMSDNDPOYNKD 21

RESULT 9

US-09-649-885-2

; Sequence 2, Application US/09649885

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Substitutes for Modified Starch and

; FILE REFERENCE: 0358D2

; CURRENT APPLICATION NUMBER: US/09/649,885

; EARLIER FILING DATE: 2000-08-28

; EARLIER APPLICATION NUMBER: US 09/008,172

; EARLIER FILING DATE: 1998-01-16

; EARLIER APPLICATION NUMBER: 08/482,711

; EARLIER FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 1430

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-649-885-2

Query Match 63.0%; Score 75; DB 20; Length 1430;

Best Local Similarity 71.4%; Pred. No. 0.034;

Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDPTYLHD 21

Db 495 AINHLSILEAMSDNDPOYNKD 515

RESULT 10

US-09-740-274-6

; Sequence 6, Application US/09740274

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Glucan-containing Compositions and Paper

; FILE REFERENCE: 0357CRD

; CURRENT APPLICATION NUMBER: US/09/740,274

; EARLIER FILING DATE: 2000-12-19

; EARLIER APPLICATION NUMBER: 09/210,361

; EARLIER FILING DATE: 1998-12-11

```

1 PRIOR APPLICATION NUMBER: 09/007,999
2 PRIOR FILING DATE: 1998-01-16
3 PRIOR APPLICATION NUMBER: 08/478,704
4 PRIOR FILING DATE: 1995-06-07
5 PRIOR APPLICATION NUMBER: 09/009,620
6 PRIOR FILING DATE: 1998-01-20
7 PRIOR APPLICATION NUMBER: 08/485,243
8 PRIOR FILING DATE: 1995-06-07
9 PRIOR APPLICATION NUMBER: 09/008,172
10 PRIOR FILING DATE: 1998-01-16
11 PRIOR APPLICATION NUMBER: 08/482,711
12 PRIOR FILING DATE: 1995-06-07
13 NUMBER OF SEQ ID NOS: 6
14 SOFTWARE: FASTSEQ for Windows Version 3.0
15 SEQ ID NO 6
16 LENGTH: 1430
17 TYPE: PRN
18 ORGANISM: streptococcus mutans
19 09-740-274-6

```

Query Match	63.0%	Score 75	DB 21	Length 1430
Best Local Similarity	71.4%	Pred. No. 0.034		
Matches 15; Conservative		1; Mismatches	5; Indels	0; Gaps

```
QY      1 ANDHLSILEAWSNDNTPYLHD 21
         | : ||||| ||||| | |
Db      495 AINHLSILEAWSNDNDPQYNKD 515
```

```

11 RESULT 11
12 US-09-290-049-13
13 Sequence 13, Application US/09290049
14 GENERAL INFORMATION:
15 APPLICANT: Smith, Daniel J.
16 TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
17 TITLE OF INVENTION: CRIES
18 FILE REFERENCE: FDC98-01p2A
19 CURRENT APPLICATION NUMBER: US/09/290,049
20 CURRENT FILING DATE: 1999-04-12
21 EARLIER APPLICATION NUMBER: 60/081,550
22 EARLIER FILING DATE: 1998-04-13
23 EARLIER APPLICATION NUMBER: 60/115,142
24 EARLIER FILING DATE: 1999-01-08
25 NUMBER OF SEQ ID NOS: 19
26 SOFTWARE: FastSeq for Windows Version 3.0
27 SEQ ID NO 13
28 LENGTH: 21
29 TYPE: PRF
30 ORGANISM: S. downei
31 US-09-290-049-13

```

Query Match	61.3%	Score 73	DB 16	Length 21
Best Local Similarity	73.7%	Pred. No. 0.00073		
Matches 14; Conservative		1; Mismatches 4;	Indels 0;	Gaps 0;

```
QY 1 ANDHLSILEAWSNDNTPYL 19
    | | | | | | | | | |
Db 1 AIDHLSILEAWSGNDNDYV 19
```

```

RESULT 12
US-09-499-203-2
; Sequence 2, Application US/09499203
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNOTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P

```

CURRENT APPLICATION NUMBER: US-09/499, 203  
 CURRENT FILING DATE: 2000-02-08  
 NUMBER OF SEQ ID NOS: 54  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 2  
 LENGTH: 2057  
 TYPE: PRT  
 ORGANISM: Leuconostoc mesenteroides  
 US-09-499-203-2

Query Match	50.4%	Score	60	DB	18	Length	2057
Best Local Similarity	55.0%	Pred. No.	11				
Matches	11	Conservative	3	Mismatches	6	Indels	0
						Gaps	0

```
QY      1 ANDHLSILEAWSNDNTPYLH 20
        || ||||| | : | ::
Db      665 ANKHLSTLEDWNGKDPQYVN 684
```

```

RESULT 13
US-09-595-298A-661
: Sequence 661, Application US/09595298A
: GENERAL INFORMATION:
: APPLICANT: ALEXANDROV, Nikolai
: APPLICANT: BROVER, Vyacheslav
: TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept
: TITLE OF INVENTION: Theraby
: FILE REFERENCE: 2750-0953p
: CURRENT APPLICATION NUMBER: US/09/595, 298A
: CURRENT FILING DATE: 2000-06-16
: NUMBER OF SEQ ID NOS: 2756
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 661
: LENGTH: 306
: TYPE: PRT
: ORGANISM: Arabidopsis Thaliana
: FEATURE:
: NAME/KEY: peptide
: LOCATION: (1)..(306)
: OTHER INFORMATION: Ceres Seq. ID no. 1026741
: NAME/KEY: misc_feature
: LOCATION: ()..()
: OTHER INFORMATION: xaa is any aa, unknown or other
US-09-595-298A-661

```

Query Match	42.0%;	Score 50;	DB 19;	Length 306;
Best Local Similarity	52.9%;	Pred. NO. 49;		
Matches 9; Conservative			5; Indels 0;	Gaps 0;
			3; Mismatches	

QY 3 DHL\$ILEAWS\$DNDTPYL 19  
| : | : | | | | | :  
Db 200 DYEQIMEAWS\$DKGTLV 216

```

RESULT 14
US-09-595-298A-660
; Sequence 660, Application US/09595298A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0953P
; CURRENT APPLICATION NUMBER: US/09/595, 298A
; CURRENT FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 2756
; SOFTWARE: PatentIn version 3.0
SEQ ID NO 660
LENGTH: 340
;
; TYPE: PRT
;
; ORGANISM: Arabidopsis Thaliana
;

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:49:47 ; Search time 87.3 Seconds

(without alignments)  
5.413 Million cell updates/sec

Title: US-09-290-049a-1

Perfect score: 119

Sequence: 1 ANDHSLTLEAMSDNDPRLHD 21

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

21 number of hits satisfying chosen parameters: 212252

Maximum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents-AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	1475	US-09-007-999-2	Sequence 2, Appli
2	119	100.0	1475	US-09-210-361-2	Sequence 2, Appli
3	110	92.4	1375	US-09-210-361-4	Sequence 4, Appli
4	75	63.0	1430	US-09-008-172-2	Sequence 2, Appli
5	75	63.0	1430	US-09-210-361-6	Sequence 6, Appli
6	63	52.9	1577	US-08-793-824-2	Sequence 6, Appli
7	45	37.8	385	US-08-361-920-23	Sequence 23, Appli
8	45	37.8	385	US-08-479-930-23	Sequence 23, Appli
9	45	37.8	385	US-08-483-432-23	Sequence 23, Appli
10	44	37.0	384	US-09-311-170-2	Sequence 2, Appli
11	43	36.1	2465	US-08-596-291-3	Sequence 3, Appli
12	43	36.1	2465	US-09-100-804-3	Sequence 3, Appli
13	43	36.1	2466	US-09-080-855-12	Sequence 12, Appli
14	43	36.1	2466	PCR-US94-09943-2	Sequence 46, Appli
15	43	36.1	2485	US-09-290-640-46	Sequence 3, Appli
16	42	35.3	259	US-08-277-231A-3	Sequence 3, Appli
17	42	35.3	259	US-08-473-750-6	Sequence 6, Appli
18	42	35.3	259	US-08-477-326-6	Sequence 6, Appli
19	42	35.3	989	US-08-070-301-16	Sequence 16, Appli
20	41	34.5	480	US-08-272-255-12	Sequence 12, Appli
21	41	34.5	480	PCR-US95-08565-12	Sequence 12, Appli
22	40	33.6	196	US-08-933-750C-35	Sequence 35, Appli
23	40	33.6	196	US-09-234-613-35	Sequence 35, Appli
24	40	33.6	307	US-08-332-562A-83	Sequence 83, Appli
25	40	33.6	334	US-09-362-473-6	Sequence 6, Appli
26	40	33.6	455	US-09-362-473-6	Sequence 6, Appli
27	40	33.6	761	US-09-625-188-14	Sequence 14, Appli

28	40	33.6	839	4	US-09-197-636-2	Sequence 2, Appli
29	40	33.6	839	4	US-09-197-636-4	Sequence 4, Appli
30	40	33.6	839	4	US-09-197-636-8	Sequence 8, Appli
31	40	33.6	2860	2	US-08-826-267-2	Sequence 2, Appli
32	40	33.6	3898	4	US-08-750-717-2	Sequence 2, Appli
33	39.5	33.2	337	3	US-09-032-372-2	Sequence 2, Appli
34	39.5	33.2	500	6	5171684-2	Patent No. 5171684
35	39.5	33.2	567	1	US-08-258-261B-4	Sequence 4, Appli
36	39.5	33.2	567	1	US-08-456-837-4	Sequence 4, Appli
37	39.5	33.2	567	1	US-08-457-342-4	Sequence 4, Appli
38	39.5	33.2	567	1	US-08-457-646A-4	Sequence 4, Appli
39	39.5	33.2	567	1	US-08-458-076A-4	Sequence 4, Appli
40	39.5	33.2	567	1	US-08-457-335A-4	Sequence 4, Appli
41	39.5	33.2	567	2	US-08-729-214-4	Sequence 4, Appli
42	39.5	33.2	567	3	US-09-028-934-4	Sequence 4, Appli
43	39	32.8	755	3	US-09-071-101-2	Sequence 2, Appli
44	39	32.8	755	3	US-09-369-618-2	Sequence 2, Appli
45	39	32.8	755	3	US-09-369-617-2	Sequence 2, Appli

#### ALIGNMENTS

```
RESULT 1
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 2
; LENGTH: 1475
; TYPE: PR1
; ORGANISM: S. Typhimurium
US-09-007-999-2

Query Match 100.0%; Score 119; DB 3; Length 1475;
Best Local Similarity 100.0%; Pred. No. 6,4e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHSLTLEAMSDNDPRLHD 21
DB 481 ANDHSLTLEAMSDNDPRLHD 501

RESULT 2
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 5284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
```

```
EARLIER APPLICATION NUMBER: 09/008,172
EARLIER FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/482,711
EARLIER FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1475
TYPE: PRT
ORGANISM: Streptococcus mutans
US-09-210-361-2
```

```
Query Match          100.0%; Score 119; DB 4; Length 1475;
Best Local Similarity 100.0%; Pred. No. 6,4e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ANDHLSLEAWSNDPPLYLD 21
|:|||||
Db 481 ANDHLSLEAWSNDPPLYLD 501
```

```
RESULT 3
US-09-210-361-4
Sequence 4, Application US/09210361
Patent No. 6284479
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Substitutes for Modified Starches and
FILE REFERENCE: 0357CR
CURRENT APPLICATION NUMBER: US/09/210,361
EARLIER FILING DATE: 1998-12-11
EARLIER APPLICATION NUMBER: 09/007,999
EARLIER FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/478,704
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 09/009,620
EARLIER FILING DATE: 1998-01-20
EARLIER APPLICATION NUMBER: 08/485,243
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 09/008,172
EARLIER FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/482,711
EARLIER FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 1375
TYPE: PRT
ORGANISM: streptococcus mutans
US-09-210-361-4
```

```
Query Match          92.4%; Score 110; DB 4; Length 1375;
Best Local Similarity 95.2%; Pred. No. 1.7e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 ANDHLSLEAWSNDPPLYLD 21
|:|||||
Db 507 ANDHLSLEAWSNDPPLYLD 527
```

```
RESULT 4
US-09-008-172-2
Sequence 2, Application US/09008172
Patent No. 6127602
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Substitutes for Modified Starch and
FILE REFERENCE: 0358D
CURRENT APPLICATION NUMBER: US/09/008,172
```

```
CURRENT FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/482,711
EARLIER FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1430
TYPE: PRT
ORGANISM: Streptococcus mutans
US-09-008-172-2
```

```
Query Match          63.0%; Score 75; DB 3; Length 1430;
Best Local Similarity 71.4%; Pred. No. 0.00084;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 ANDHLSLEAWSNDPPLYLD 21
|:|||||
Db 495 AINHLSTLEAWSNDPQYKND 515
```

```
RESULT 5
US-09-210-361-6
Sequence 6, Application US/09210361
Patent No. 6284479
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Substitutes for Modified Starches and
FILE REFERENCE: 0357CR
CURRENT APPLICATION NUMBER: US/09/210,361
EARLIER FILING DATE: 1998-12-11
EARLIER APPLICATION NUMBER: 09/007,999
EARLIER FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/478,704
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 09/009,620
EARLIER FILING DATE: 1998-01-20
EARLIER APPLICATION NUMBER: 08/485,243
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 09/008,172
EARLIER FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/482,711
EARLIER FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6
LENGTH: 1430
TYPE: PRT
ORGANISM: streptococcus mutans
US-09-210-361-6
```

```
Query Match          63.0%; Score 75; DB 4; Length 1430;
Best Local Similarity 71.4%; Pred. No. 0.00084;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 ANDHLSLEAWSNDPPLYLD 21
|:|||||
Db 495 AINHLSTLEAWSNDPQYKND 515
```

```
RESULT 6
US-08-793-824-2
Sequence 2, Application US/08793824
Patent No. 5981838
GENERAL INFORMATION:
APPLICANT: Simpson, Christine Lynn
APPLICANT: Giffard, Philip Morrison
APPLICANT: Jacques, Nicholas Anthony
TITLE OF INVENTION: Genetic Manipulation of Plants to
INCREASE STORED CARBOHYDRATES
NUMBER OF SEQUENCES: 2
```



1 CORRESPONDENCE ADDRESS:  
2 ADDRESS: Griffith Hack & Co  
3 STREET: Level 8, 168 Walker Street  
4 CITY: NO. 5961838th Sydney  
5 STATE: New South Wales  
6 COUNTRY: Australia  
7 ZIP: 2060  
8  
9 COMPUTER READABLE FORM:  
10 MEDIUM TYPE: Floppy disk  
11 COMPUTER: IBM PC compatible  
12 OPERATING SYSTEM: PC-DOS/MS-DOS  
13 SOFTWARE: Patentin Release #1.0, Version #1.30  
14  
15 CURRENT APPLICATION DATA:  
16 APPLICATION NUMBER: US/08/793,824  
17 FILING DATE:  
18 CLASSIFICATION: 800  
19 PRIOR APPLICATION DATA:  
20 APPLICATION NUMBER: AU PM7643  
21 FILING DATE: 24-AUG-1994  
22 TELECOMMUNICATION INFORMATION:  
23 TELEPHONE: 61 2 9957 5944  
24 TELEFAX: 61 2 957 6286  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801

Query Match	52.9%	Score 63	DB 2	Length 1577	
Best Local Similarity	72.2%	Pred. No. 0.084			
Matches 13, Conservative	0	Mismatches 5	Indels 0	Gaps 0	

```
QY      4 HLSILEAWSNDTPLYHD 21
          |||||..|||
Db      591 HLSILEAWSYNDHQYKND 608
```

7  
7 RESULT 7  
7 US-08-361-920-23  
7 Sequence 23. Application US/08361920  
7 Patent No. 5457046  
7  
7 GENERAL INFORMATION:  
7 APPLICANT: Woeldike, Helle F.  
7 APPLICANT: Hagen, Frederick  
7 APPLICANT: Hjort, Carsten M.  
7 APPLICANT: Sven, Hastrup  
7  
7 TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
7  
7 TITLE OF INVENTION: or Hemicellulose  
7  
7 NUMBER OF SEQUENCES: 85  
7  
7 CORRESPONDENCE ADDRESS:  
7 ADDRESSEE: No. 54570460 No. 5457046disk of No. 5457046th America, Inc  
7 STREET: 405 Lexington Avenue, 62nd Floor  
7  
7 CITY: New York  
7  
7 STATE: New York  
7  
7 COUNTRY: United States of America  
7  
7 ZIP: 10174-6201  
7  
7 COMPUTER READABLE FORM:  
7  
7 MEDIUM TYPE: Floppy disk  
7  
7 COMPUTER: IBM PC compatible  
7  
7 OPERATING SYSTEM: PC-DOS/MS-DOS  
7  
7 SOFTWARE: PatentIn Release #1.0, Version #1.25  
7  
7 CURRENT APPLICATION DATA:  
7  
7 APPLICATION NUMBER: US/08/361, 920  
7  
7 FILING DATE:  
7  
7 CLASSIFICATION: 435  
7  
7 PRIOR APPLICATION DATA:  
7

```

      APPLICATION NUMBER: US 07/940,860
      FILING DATE: 28-OCT-1992
      APPLICATION NUMBER: DK 1158/90
      FILING DATE: 09-MAY-1990
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/DK91/00124
      FILING DATE: 08-MAY-1991
      ATTORNEY/AGENT INFORMATION:
      NAME: Lambiris, Elias J.
      REGISTRATION NUMBER: 33,728
      REFERENCE/DOCKET NUMBER: 3435, 204-US
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 212-867-0123
      TELEFAX: 212-867-0298
      INFORMATION FOR SEQ ID NO: 23:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 385 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      US-08-361-920-2?

Query Match      37.8%      Score 45;  DB 1;  Length 385;
Best Local Similarity 24.4%;  Pred. No. 13;
Matches 10;  Conservative 5;  Mismatches 6;  Indels 20;  Gaps 1.

Oy      1  ANDHLSLEA-----WSDNDTPYLD 21
      |||:|:::|      |  |  |  |  |
Db      329  ANDYLLTVMNACTLAVPKCVGITYWGVSDKDSWRPGNDPLLYD 369

```

```

QY      1 ANDHLSLEA-----WSDNTPYLHD 21
      |||:::|          | | | |
DB      329 ANDYLIVNACLAIPKCVGITVWGVSDKDSMRPGDNPILD 365

```

RESULT 8  
 US-08-479-939-23  
 Sequence 23 Application US/08479939  
 Patent No. 5686593  
 GENERAL INFORMATION:  
 APPLICANT: Woelldike, Helle F.  
 APPLICANT: Hagen, Frederick  
 APPLICANT: Hjort, Carsten M.  
 APPLICANT: Sven, Hastrup  
 TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
 TITLE OF INVENTION: or Hemicellulose  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: No. 56865930 No. 5686593disk of No. 5686593th America, Inc.  
 STREET: 405 Lexington Avenue, 62nd Floor  
 CITY: New York  
 STATE: New York  
 COUNTRY: United States of America  
 ZIP: 10174-6201  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/479,939  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/361,920  
 FILING DATE: 22-DEC-1994  
 APPLICATION NUMBER: US 07/940,860  
 FILING DATE: 28-OCT-1992  
 APPLICATION NUMBER: DK 1158/90  
 FILING DATE: 09-MAY-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/OK91/00124  
 FILING DATE: 08-MAY-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lambdilis, Elias J.  
 REGISTRATION NUMBER: 33,728



NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31.616  
REFERENCE/DOCKET NUMBER: L0461/7000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
TELEX: 92-1742 EZEKIL  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2465 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-596-291-3

Query Match 36.1%; Score 43; DB 2; Length 2465;  
Best Local Similarity 56.2%; Pred. No. 2.5e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

4 H1SILE--AMSDNDTP 17  
1:1 111111  
2347 H1SHLFTAMPDHDTP 2362

## RESULT 12

US-09-100-804-3  
Sequence 3, Application US/09100804  
Patent No. 6066472

GENERAL INFORMATION:  
APPLICANT: GONZ, LEONEL JORGE  
APPLICANT: SARAS, JAN  
APPLICANT: CLAESSON-WEISH, LENA  
APPLICANT: HELDIN, CARL-HENRIK  
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSER: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/100.804

## FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/596,291

FILING DATE: 09-AUG-1996  
APPLICATION NUMBER: US 08/115,573

FILING DATE: 01-SEP-1993  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/09943  
FILING DATE: 01-SEP-1994

ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31.616  
REFERENCE/DOCKET NUMBER: L0461/7003

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441  
TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2465 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-100-804-3

Query Match 36.1%; Score 43; DB 3; Length 2465;  
Best Local Similarity 56.2%; Pred. No. 2.5e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 4 H1SILE--AMSDNDTP 17  
1:1 111111  
Db 2347 H1SHLFTAMPDHDTP 2362

## RESULT 13

US-09-080-855-12  
Sequence 12, Application US/09080855A  
Patent No. 6083721

GENERAL INFORMATION:  
APPLICANT: Saras, Jan  
APPLICANT: Eranza, Petra  
APPLICANT: Aspenstrm, Pontus  
APPLICANT: Hellman, Ulf  
APPLICANT: Gonzalez, Leonel Jorge  
APPLICANT: Heldin, Carl-Henrik  
TITLE OF INVENTION: PARG, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPL1  
FILE REFERENCE: L0461/7030  
CURRENT APPLICATION NUMBER: US/09/080,855A  
CURRENT FILING DATE: 1998-05-18  
EARLIER APPLICATION NUMBER: 08/805,583  
EARLIER FILING DATE: 1997-02-25  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 2466  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-080-855-12

Query Match 36.1%; Score 43; DB 3; Length 2466;  
Best Local Similarity 56.2%; Pred. No. 2.5e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 4 H1SILE--AMSDNDTP 17  
1:1 111111  
Db 2348 H1SHLFTAMPDHDTP 2363

## RESULT 14

PCT-US94-09943-2  
Sequence 2, Application PC/TUS9409943

GENERAL INFORMATION:

APPLICANT:

APPLICANT:

APPLICANT:

APPLICANT:

APPLICANT:

APPLICANT:

APPLICANT:

APPLICANT:

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/09943
: FILING DATE: 01-SEP-1994
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/115,573
: FILING DATE: 01-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: TWOMEY, MICHAEL J.
: REGISTRATION NUMBER: P-38,349
: REFERENCE/DOCKET NUMBER: L0461/7000WO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/720-3500
: TELEFAX: 617/720-2441
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2466 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US94-09943-2

```

```

Query Match          36.1%; Score 43; DB 5; Length 2466;
Best Local Similarity 56.2%; Pred. No. 2.5e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 4 HLSILE--AWSNDTP 17
|:|:| 111:111
DB 2348 HISHLNTAWPDHDP 2363

```

```

RESULT 15
US-09-290-640-46
: Sequence 46, Application US/09290640
: Patent No. 6204055
: GENERAL INFORMATION:
: APPLICANT: Dean, Nicholas M.
: APPLICANT: Marcussen, Eric G.
: TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling
: FILE REFERENCE: ISPH-0351
: CURRENT APPLICATION NUMBER: US/09/290,640
: CURRENT FILING DATE: 1999-04-12
: NUMBER OF SEQ ID NOS: 85
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 46
: LENGTH: 2485
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-290-640-46

```

```

Query Match          36.1%; Score 43; DB 4; Length 2485;
Best Local Similarity 56.2%; Pred. No. 2.5e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 4 HLSILE--AWSNDTP 17
|:|:| 111:111
DB 2367 HISHLNTAWPDHDP 2382

```

Search completed: March 27, 2002, 13:59:26  
Job time: 579 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:54:33 ; Search time 137.48 seconds  
(without alignments)  
10.540 Million cell updates/sec

Title: US-09-290-049A-1

Perfect score: 119

Sequence: 1 ANDHLSILEAWSNDNPTPLYHD 21

Scoring table: BLOSUM62

Searched: 332938 seqs, 6899538 residues

Minimum DB seg length: 0  
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCr\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep1:\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep1:\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	21	US-09-562-328-20	Sequence 20, Appl
2	119	100.0	21	US-09-290-049A-1	Sequence 1, Appl
3	119	100.0	21	US-09-290-049A-10	Sequence 10, Appl
4	110	92.4	21	US-09-562-328-21	Sequence 21, Appl
5	108	90.8	21	US-09-562-328-23	Sequence 23, Appl
6	108	90.8	21	US-09-562-328-25	Sequence 25, Appl
7	108	90.8	21	US-09-290-049A-12	Sequence 12, Appl
8	108	90.8	21	US-09-290-049A-14	Sequence 14, Appl
9	75	63.0	21	US-09-562-328-22	Sequence 22, Appl
10	75	63.0	21	US-09-290-049A-11	Sequence 11, Appl
11	75	63.0	21	US-09-604-957-4	Sequence 4, Appl
12	73	61.3	21	US-09-562-328-24	Sequence 24, Appl
13	73	61.3	21	US-09-290-049A-13	Sequence 13, Appl
14	71	59.7	523	US-09-604-957-5	Sequence 5, Appl
15	60	50.4	584	US-09-604-957-6	Sequence 6, Appl
16	49.5	41.6	423	US-09-708-427-55326	Sequence 55326, A
17	49.5	41.6	511	US-09-708-427-55325	Sequence 55325, A
18	49.5	41.6	549	US-09-708-427-55324	Sequence 55324, A
19	49	41.2	302	PCT-US02-03987-14004	Sequence 14004, A
20	49	41.2	302	US-09-815-242-14004	Sequence 14004, A
21	49	41.2	302	US-10-072-851-14004	Sequence 14004, A
22	49	41.2	535	US-09-604-957-7	Sequence 7, Appl
23	49	41.2	1278	US-09-604-957-3	Sequence 3, Appl
24	47.5	39.9	434	US-09-708-427-7994	Sequence 7994, Ap
25	47.5	39.9	519	US-09-708-427-7993	Sequence 7993, Ap

26	47	39.5	15	7	US-10-067-649-90	Sequence 90, Appl
27	47	39.5	366	6	US-09-708-427-27415	Sequence 27415, A
28	47	39.5	367	6	US-09-708-427-27415	Sequence 27415, A
29	47	39.5	378	6	US-09-708-427-27414	Sequence 27414, A
30	46.5	39.1	1604	6	US-09-888-615-73	Sequence 73, Appl
31	46	38.7	199	6	US-09-605-703B-2226	Sequence 2226, Ap
32	46	38.7	2590	6	US-09-897-516-7203	Sequence 7203, Ap
33	46	38.7	2603	6	US-09-897-516-5891	Sequence 5891, Ap
34	45	37.8	185	6	US-09-897-516-5719	Sequence 5719, Ap
35	45	37.8	236	6	US-09-154-750A-80	Sequence 80, Appl
36	45	37.8	638	6	US-09-708-427-25322	Sequence 25322, A
37	45	37.8	765	6	US-09-708-427-25321	Sequence 25321, A
38	45	37.8	947	6	US-09-708-427-25320	Sequence 25320, A
39	44.5	37.4	499	6	US-09-614-150-42525	Sequence 42525, A
40	44	37.0	322	1	PCT-US02-03987-11634	Sequence 11634, A
41	44	37.0	322	1	US-09-815-242-11634	Sequence 11634, A
42	44	37.0	322	7	US-10-072-851-11634	Sequence 11634, A
43	44	37.0	331	1	PCT-US02-03987-10188	Sequence 10188, A
44	44	37.0	331	6	US-09-815-242-10188	Sequence 10188, A
45	44	37.0	331	7	US-10-072-851-10188	Sequence 10188, A

## ALIGNMENTS

RESULT 1  
US-09-562-328-20  
Sequence 20, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT FILING DATE: 2000-05-01  
CURRENT APPLICATION NUMBER: US/09/562.328  
PRIOR FILING DATE: 1999-04-09  
PRIOR APPLICATION NUMBER: 09/288,965  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 20  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-20

Query Match 100.0%; Score 119; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2,9e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ANDHLSILEAWSNDNPTPLYHD 21  
DB 1 ANDHLSILEAWSNDNPTPLYHD 21  
RESULT 2  
US-09-290-049A-1  
Sequence 1, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: 1564.1008-002  
CURRENT APPLICATION NUMBER: US/09/290.049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0

```

; SEQ ID NO 1
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EAW peptide
US-09-290-049A-1
```

```

Query Match          100.0%; Score 119; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 2,9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 ANDHLSILEAWSNDPTPYLHD 21
    |||
Db 1 ANDHLSILEAWSNDPTPYLHD 21
```

```

ULT 3
09-290-049A-10
; Sequence 10, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049A-10
```

```

Query Match          100.0%; Score 119; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 2,9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
1 ANDHLSILEAWSNDPTPYLHD 21
  |||
1 ANDHLSILEAWSNDPTPYLHD 21
```

```

RESULT 4
US-09-562-328-21
; Sequence 21, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-21
```

```

Query Match          92.4%; Score 110; DB 6; Length 21;
Best Local Similarity 95.2%; Pred. No. 7,4e-11;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 1 ANDHLSILEAWSNDPTPYLHD 21
    |||
Db 1 ANDHLSILEAWSNDPTPYLHD 21
```

```

RESULT 5
US-09-562-328-23
; Sequence 23, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-23
```

```

Query Match          90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1,5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 ANDHLSILEAWSNDPTPYLHD 21
    |||
Db 1 ANNHVSIVEAWSNDPTPYLHD 21
```

```

RESULT 6
US-09-562-328-25
; Sequence 25, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-25
```

```

Query Match          90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1,5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 ANDHLSILEAWSNDPTPYLHD 21
    |||
Db 1 ANNHVSIVEAWSNDPTPYLHD 21
```

```

RESULT 7
US-09-290-049A-12
```

```
; Sequence 12, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049A-12
```

```
Query Match          90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSLEAWSNDNDTPYLHD 21
Db      1 ANNHVSIVEAWSNDNDTPYLHD 21
```

```
RESULT      8
US-09-290-049A-14
; Sequence 14, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. sobrinus
US-09-290-049A-14
```

```
Query Match          90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSLEAWSNDNDTPYLHD 21
Db      1 ANNHVSIVEAWSNDNDTPYLHD 21
```

```
RESULT      9
US-09-562-328-22
; Sequence 22, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
```

```
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-22
```

```
Query Match          63.0%; Score 75; DB 6; Length 21;
Best Local Similarity 71.4%; Pred. No. 2.3e-05;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSLEAWSNDNDTPYLHD 21
Db      1 AINHLSLEAWSNDNDPQYNKD 21
```

```
RESULT      10
US-09-290-049A-11
; Sequence 11, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049A-11
```

```
Query Match          63.0%; Score 75; DB 6; Length 21;
Best Local Similarity 71.4%; Pred. No. 2.3e-05;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSLEAWSNDNDTPYLHD 21
Db      1 AINHLSLEAWSNDNDPQYNKD 21
```

```
RESULT      11
US-09-604-957-4
; Sequence 4, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUTI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
```

TYPE: PRF  
ORGANISM: Streptococcus mutans  
US-09-604-957-4

Query Match 63.0%; Score 75; DB 6; Length 545;  
Best Local Similarity 71.4%; Pred. No. 0.00094;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDTPYL 21  
DB 75 ANHLSLEAMSDNDPQYND 95

RESULT 12  
US-09-562-328-24  
Sequence 24, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.

APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 24  
LENGTH: 21  
TYPE: PRF  
ORGANISM: Streptococcus sp.  
US-09-562-328-24

Query Match 61.3%; Score 73; DB 6; Length 21;  
Best Local Similarity 73.7%; Pred. No. 4.7e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDTPYL 19  
DB 1 AIDHLSLEAMSGNDNDYV 19

RESULT 13  
US-09-290-049A-13  
Sequence 13, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: 1564.1008-002  
CURRENT APPLICATION NUMBER: US/09/290,049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 21  
TYPE: PRF  
ORGANISM: S. downei  
US-09-290-049A-13

Query Match 61.3%; Score 73; DB 6; Length 21;  
Best Local Similarity 73.7%; Pred. No. 4.7e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDTPYL 19  
DB 1 AIDHLSLEAMSGNDNDYV 19

RESULT 14  
US-09-604-957-5  
Sequence 5, Application US/09604957  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIJKHUIZEN, LOBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604,957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 5  
LENGTH: 523  
TYPE: PRF  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Query Match 59.7%; Score 71; DB 6; Length 523;  
Best Local Similarity 66.7%; Pred. No. 0.0038;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDTPYL 21  
DB 75 ANHLSLEAMSHNDPLVYTD 95

RESULT 15  
US-09-604-957-6  
Sequence 6, Application US/09604957  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIJKHUIZEN, LOBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604,957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 6  
LENGTH: 584  
TYPE: PRF  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-6

Query Match 50.4%; Score 60; DB 6; Length 584;  
Best Local Similarity 55.0%; Pred. No. 0.23;  
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDTPYL 20  
DB 75 ANHLSLEAMWGKDPQYVN 94

Search completed: March 27, 2002, 14:22:45  
Job time: 1692 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:53:18 ; Search time 102.51 Seconds  
(without alignments)  
15.605 Million cell updates/sec

Title: US-09-290-049a-1

Perfect score: 119

Sequence: 1 ANDHLSILEAWSNDNPTPLHD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	1475	2 B33135	gtfB protein precu
2	110	92.4	1375	2 JT0345	dextranucrase (EC
3	108	90.8	1592	2 A38175	glucosyltransferas
4	78	65.5	1365	2 A41463	glucosyltransferas
5	75	63.0	1431	2 A45866	dextranucrase (EC
6	74	62.2	1508	2 T31098	probable dextranu
7	66	55.5	1449	2 T30857	glucosyltransferas
8	66	55.5	1449	2 T30552	glucosyltransferas
9	63	52.9	1577	2 T30858	glucosyltransferas
10	58	48.7	1518	2 A44811	glucosyltransferas
11	57	47.9	1331	2 B48445	glyceraledehyde-3-p
12	57	47.9	1599	2 S22737	glucosyltransferas
13	52	43.7	4848	2 T30289	pristinamycin I sy
14	50	42.0	347	2 T48610	hypothetical prote
15	49.5	41.6	349	2 T03963	rif2 nuclear restor
16	49	41.2	336	1 Q0BE40	BGLF2 protein - hu
17	48	40.3	490	2 H70538	probable ppdk prot
18	47.5	39.9	331	1 DEUTGC	glyceraledehyde-3-p
19	47.5	39.9	519	2 C86372	hypothetical prote
20	47	39.5	175	2 C86205	hypothetical prote
21	47	39.5	378	2 T04254	hypothetical prote
22	47	39.5	524	2 D82220	conserved hypotet
23	47	39.5	525	2 T40088	RhogeF domain cont
24	46	38.7	275	2 S55978	hypothetical prote
25	46	38.7	335	2 E86568	glyceraledehyde-3-p
26	46	38.7	335	2 B22053	glyceraledehyde 3-p
27	46	38.7	476	1 S58229	salicylate biosynt
28	45.5	38.2	96	2 E81786	conserved hypotet
29	45	37.8	236	2 S62732	guanidinacetate N

30	45	37.8	947	2 E86362	hypothetical prote
31	45	37.8	1159	2 T43461	probable phosphodi
32	44.5	37.4	210	2 H83332	conserved hypotet
33	44	37.0	247	2 P00178	glyceraledehyde-3-p
34	44	37.0	331	1 DEECG3	glyceraledehyde-3-p
35	44	37.0	331	2 D85788	glyceraledehyde-3-p
36	44	37.0	333	1 DECHG3	glyceraledehyde-3-p
37	44	37.0	333	2 JC5370	glyceraledehyde-3-p
38	44	37.0	337	1 DEJUCG	glyceraledehyde-3-p
39	44	37.0	337	1 DEZMGC	glyceraledehyde-3-p
40	44	37.0	337	1 DEBHG	glyceraledehyde-3-p
41	44	37.0	436	2 T24953	hypothetical prote
42	44	37.0	468	2 S61964	probable membrane
43	44	37.0	1208	2 B82091	exodexyribonuclea
44	44	37.0	1313	2 T29027	hypothetical prote
45	43.5	36.6	97	2 S12958	tachykinin delta p

## ALIGNMENTS

```

RESULT 1
B33135
gtfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 15-Oct-1999
C:Accession: B33135; A33128
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013
A:Accession: B33135
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SH1>
A:Cross-references: GB:M17361; NID:g153639; PID:AAA8588.1; PID:g153640
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
submitted to the Protein Sequence Database, September 1990
A:Reference number: A33128
A:Accession: A33128
A>Status: Preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-171, 173-641, 'N', 643-1475 <SH2>
A:Experimental source: strain GS-5
C:Superfamily: cpl repeat homology
F:1096-1115/Domain: cpl repeat homology <CP1>
F:1124-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>
F:1354-1373/Domain: cpl repeat homology <CP4>
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 100.0%; Score 119; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 3.9e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0;

QY 1 ANDHLSILEAWSNDNPTPLHD 21
DB 481 ANDHLSILEAWSNDNPTPLHD 501

RESULT 2
JT0345
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999
C:Accession: JT0345; C33135
R:Ueda, S.; Shiroza, T.; Kuramitsu, H. K.
Gene 69, 101-109, 1988
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
A:Reference number: JT0345; MUID:89137980
A:Accession: JT0345

```

A:Molecule type: DNA  
A:Residues: 1-1375 <UED>  
C:Accession: A41483  
R:Gillmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
J. Infect. Immun. 58, 2452-2458, 1990  
A:Title: Analysis of the Streptococcus downei gtfB gene, which specifies a glucosyltransferase  
A:Reference number: A41483; MUID:90316665  
A:Accession: A41483  
A:Molecule type: DNA  
A:Residues: 1-349 <SHL>  
A:Cross-References: GB:M17361  
C:Genetics:  
A:Gene: gtfC  
C:Function:  
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
C:Superfamily: cpl repeat homology  
C:Keywords: duplication; glucosyltransferase; hexosyltransferase  
-34/Domain: signal sequence #status predicted <SIG>  
-35-1375/Product: glucosyltransferase #status predicted <MAT>  
F:1126-1145/Domain: cpl repeat homology <CP1>  
F:1253-1272/Domain: cpl repeat homology <CP2>  
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 92.4%; Score 110; DB 2; Length 1375;  
Best Local Similarity 95.2%; Pred. No. 9, 1e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDTPYLHD 21  
DB 507 ANDHLSILEAWSNDTPYLHD 527

RESULT 3  
A38175  
glucosyltransferase precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C>Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
A:Accession: A38175  
R:Abdo, H.; Msumura, T.; Kodana, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus  
A:Reference number: A38175; MUID:91123227  
A:Accession: A38175  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1592 <ABO>  
C:Cross-References: GB:D90213; NID:9217032; PIDN:BAAL4241.1; PID:d1014946; PID:9217033  
C:Superfamily: cpl repeat homology  
F:1093-1112/Domain: cpl repeat homology <CP1>  
F:1222-1241/Domain: cpl repeat homology <CP2>  
F:1287-1306/Domain: cpl repeat homology <CP3>  
F:1330-1351/Domain: cpl repeat homology <CP4>  
F:1352-1371/Domain: cpl repeat homology <CP5>  
F:1402-1420/Domain: cpl repeat homology <CP6>  
F:1465-1484/Domain: cpl repeat homology <CP7>  
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 90.8%; Score 108; DB 2; Length 1592;  
Best Local Similarity 85.7%; Pred. No. 2, 2e-08;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDTPYLHD 21  
DB 477 ANNHVSIVEAWSNDTPYLHD 497

RESULT 4  
A41483  
glucosyltransferase (EC 2.4.1.-) gtfB precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C:Accession: A41483  
R:Gillmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
J. Infect. Immun. 58, 2452-2458, 1990  
A:Title: Analysis of the Streptococcus downei gtfB gene, which specifies a glucosyltransferase  
A:Reference number: A41483; MUID:90316665  
A:Accession: A41483  
A:Molecule type: DNA  
A:Residues: 1-1365 <GIL>  
A:Cross-References: GB:M30943; NID:9133652; PIDN:AAA26898.1; PID:9133653  
C:Genetics:  
A:Gene: gtfB  
C:Superfamily: cpl repeat homology  
C:Keywords: glucosyltransferase; hexosyltransferase

Query Match 65.5%; Score 78; DB 2; Length 1365;  
Best Local Similarity 71.4%; Pred. No. 0.00087;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDTPYLHD 21  
DB 467 AIDHLSILEAWSNDTPYLHD 487

RESULT 5  
A45866  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans  
C:Species: Streptococcus mutans  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
A:Accession: A45866  
R:Honda, O.; Kato, C.; Kuramitsu, H.K.  
J. Gen. Microbiol. 136, 2099-2105, 1990  
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the gluco  
A:Reference number: A45866; MUID:91100958  
A:Accession: A45866  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1431 <HON>  
A:Cross-References: GB:M29296  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:181-201/Domain: cpl repeat homology <CP1>  
F:1127-1146/Domain: cpl repeat homology <CP2>  
F:1192-1211/Domain: cpl repeat homology <CP3>  
F:1257-1276/Domain: cpl repeat homology <CP4>  
F:1277-1297/Domain: cpl repeat homology <CP5>  
F:1321-1340/Domain: cpl repeat homology <CP6>  
F:1341-1361/Domain: cpl repeat homology <CP7>  
F:1385-1404/Domain: cpl repeat homology <CP8>

Query Match 63.0%; Score 75; DB 2; Length 1431;  
Best Local Similarity 71.4%; Pred. No. 0.0027;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDTPYLHD 21  
DB 495 AINHLSILEAWSNDTPYLHD 515

RESULT 6  
T31098  
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides  
C:Species: Leuconostoc mesenteroides  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
A:Accession: T31098  
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
FEMS Microbiol. Lett. 159, 307-315, 1998  
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase  
A:Reference number: Z20981; MUID:98164374  
A:Accession: T31098  
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-1508 <MON>  
A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1  
A:Experimental source: strain NRRL B-1293  
C:Genetics:  
A:Gene: dsrB  
C:Function:  
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 62.2%; Score 74; DB 2; Length 1508;  
Best Local Similarity 66.7%; Pred. No. 0.0041;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 ANDHLSLEAMSDNDTPYLHD 21  
||| ||||| ||| ||| |||  
Db 563 ANOHLSTLEDMWSHNDPEYVKD 583

RESULT 7  
glycosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30857  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pr  
A:Reference number: 220909; MUID:95122197  
A:Accession: T30857  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <SIM>  
A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AAC41412.1  
C:Genetics:  
A:Gene: gltI

Query Match 55.5%; Score 66; DB 2; Length 1449;  
Best Local Similarity 72.2%; Pred. No. 0.069;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 HLSLEAMSDNDTPYLHD 21  
||| ||||| ||| ||| |||  
Db 539 HLSLEAMSHNDAYNED 556

RESULT 8  
glycosyltransferase N - Streptococcus salivarius (fragment)

C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30552  
R:Jafer, R.I.  
submitted to the EMBL Data Library, February 1998  
A:Description: Streptococcus salivarius VI477 gltN.  
A:Reference number: 220854  
A:Accession: T30552  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <JAF>  
A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1  
C:Genetics:  
A:Gene: gltN

Query Match 55.5%; Score 66; DB 2; Length 1449;  
Best Local Similarity 72.2%; Pred. No. 0.069;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
OY 4 HLSLEAMSDNDTPYLHD 21  
||| ||||| ||| ||| |||

Db 539 HLSLEAMSHNDAYNED 556

RESULT 9  
T30858

glycosyltransferase - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30858  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for  
A:Reference number: 220909; MUID:95122197  
A:Accession: T30858  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1577 <SIM>  
A:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AAC41413.1  
C:Genetics:  
A:Gene: gltM

Query Match 52.9%; Score 63; DB 2; Length 1577;  
Best Local Similarity 72.2%; Pred. No. 0.23;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 HLSLEAMSDNDTPYLHD 21  
||| ||||| ||| ||| |||  
Db 591 HLSLEAMSYNDHOYKND 608

RESULT 10

glycosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999  
C:Accession: A44811; S22726; S28809  
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase  
A:Reference number: A44811; MUID:92148377  
A:Accession: A44811  
A:Molecule type: DNA  
A:Residues: 1-1518 <GIF>  
A:Cross-references: EMBL:Z11873; NID:g47526; PID:CAAT7900.1; PID:g47527  
A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBI:81052)  
C:Genetics:  
A:Gene: gltJ  
C:Superfamily: C2\_repeat\_homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1307-1326/Dom1:n; CPl repeat homology <CP4>

Query Match 48.7%; Score 58; DB 2; Length 1518;  
Best Local Similarity 66.7%; Pred. No. 1.3;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 4 HLSLEAMSDNDTPYLHD 18  
||| ||||| ||| ||| |||  
Db 504 HSLVLEAMSLNDNHX 518

RESULT 11

B48445  
glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) - Leishmania mexicana  
C:Species: Leishmania mexicana  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 11-Jun-1999  
C:Accession: B48445; S25142  
R:Hannaert, V.; Blaauw, M.; Kohl, E.; Allert, S.; Opperdoes, F.R.; Michels, P.A.M.  
Mol. Biochem. Parasitol. 55, 115-126, 1992  
A:Title: Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-phosphat  
A:Reference number: A48445; MUID:93063042  
A:Accession: B48445

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <HAB>  
A:Cross-references: EMBL:X65220; NID:g9552; PIDD:CAA46323.1; PID:g9553  
C:Superfamily: glyceroldehyde-3-phosphate dehydrogenase  
C:Keywords: oxidoreductase

Query Match 47.9%; Score 57; DB 2; Length 331;  
Best Local Similarity 52.6%; Pred. No. 0.31;  
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 NDHLSLEAWSNDNDTPYLH 20  
||| | : | | | | |  
Db 301 NDHFWKLVSWYDNETGYSH 319

RESULT 12  
2737  
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius

C:Species: Streptococcus salivarius  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: S22737; S28810; B44811; S22727  
R:Jacques, N.  
Submitted to the EMBL Data Library, March 1992

A:Reference number: S22726  
A:Accession: S22737  
A:Molecule type: DNA

A:Residues: 1-1599 <JMC>  
A:Cross-references: EMBL:Z11872; NID:g47530; PIDD:CAA7898.1; PID:g47531  
A:Experimental source: ATCC 25975  
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
A:Reference number: A44811; MUID:92148377  
A:Accession: S28810  
A:Molecule type: DNA

A:Residues: 1-51 <GIF>  
A:Cross-references: EMBL:Z11873  
C:Genetics:  
A:Gene: gtfK  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 47.9%; Score 57; DB 2; Length 1599;  
Best Local Similarity 55.6%; Pred. No. 2;  
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 HLSLEAWSNDNDTPYLH 21  
:::||||| | | | |  
Db 494 NISILEANSHNDPYVNE 511

RESULT 13  
T30289  
pristinamycin I synthase 3 - Streptomyces pristinaespiralis

C:Species: Streptomyces pristinaespiralis  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 01-Dec-2000  
C:Accession: T30289  
R:de Crécy-Lagard, V.A.; Saurin, W.; Thibaut, D.; Gill, P.; Naudin, L.; Crouzet, J.; Bian  
Submitted to the EMBL Data Library, February 1997

A:Description: Streptogramin B biosynthesis in Streptomyces pristinaespiralis and Strept  
A:Reference number: Z20808  
A:Accession: T30289  
A:Molecule type: DNA

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4848 <DEC>  
A:Cross-references: EMBL:Y11548; NID:e1025755; PID:e307539; PIDD:CAA72312.1  
C:Genetics:  
A:Gene: snbDE  
C:Superfamily: acyl carrier protein homology; acetate--CoA ligase homology

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
F:511-951/Domain: acetate--CoA ligase homology <ACL1>  
F:966-1034/Domain: acyl carrier protein homology <ACP1>  
F:1563-2024/Domain: acetate--CoA ligase homology #status atypical <ACL2>  
F:2399-2467/Domain: acyl carrier protein homology <ACP2>  
F:2995-3427/Domain: acetate--CoA ligase homology <ACP3>  
F:3441-3509/Domain: acyl carrier protein homology <ACP3>  
F:4043-4492/Domain: acetate--CoA ligase homology <ACL4>  
F:4507-4575/Domain: acyl carrier protein homology <ACP4>  
F:998,2431,3473/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 43.7%; Score 52; DB 2; Length 4848;  
Best Local Similarity 47.1%; Pred. No. 44;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSLEAWSNDNDTP 17  
| : | : | : | : | : |  
Db 3972 ADDHARVLTGWMNDTDP 3988

RESULT 14  
T48610  
hypothetical protein F18022.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T48610  
R:Byran, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.;  
Submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24493  
A:Accession: T48610  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-347 <BEV>  
A:Cross-references: EMBL:DB  
A:Experimental source: cultivar Columbia; BAC clone F18022  
C:Genetics:  
A:Map position: 5  
A:Introns: 232/3; 296/3  
A:Note: F18022.160

Query Match 42.0%; Score 50; DB 2; Length 347;  
Best Local Similarity 52.9%; Pred. No. 4;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DHLSLEAWSNDNDTPYL 19  
| : | : | : | : | : |  
Db 200 DYEQIWEAWSKXTLYV 216

RESULT 15  
T03983  
rf2 nuclear restorer protein - maize

N:Alternate names: aldehyde dehydrogenase homolog  
C:Species: Zea mays (maize)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: T03983  
R:Cull, X.; Wise, R.P.; Schnable, P.S.  
Science 272, 1334-1336, 1996

A:Title: The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize.  
A:Reference number: Z15177; MUID:96243131  
A:Accession: T03983  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA

A:Cross-references: EMBL:U43082; NID:g1421729; PIDD:AAC49371.1; PID:g1421730  
A:Experimental source: strain B73  
C:Genetics:  
A:Map position: 9  
A:Note: rf2  
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology  
F:107-370/Domain: aldehyde dehydrogenase homology <ALD>

Query Match 41.6%; Score 49.5; DB 2; Length 549;  
Best Local Similarity 58.8%; Pred. No. 8.2;  
Matches 10; Conservative 1; Mismatches 5; Indels 1; Gaps 1;  
QY 2 NDHLSILEAWSNDNDTPY 18  
111:111111  
Db 147 NDELALETW-DNGKPY 162

Search completed: March 27, 2002, 14:01:15  
Job time: 477 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:26 ; Search time 87.3 Seconds  
(without alignments)  
4.898 Million cell updates/sec

Title: US-09-290-049a-2

Perfect score: 95

Sequence: 1 VPSYSFIRAHSEVODLIA 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCFUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95	100.0	1475	3 US-09-007-999-2	Sequence 2, Appli
2	95	100.0	1475	4 US-09-210-361-2	Sequence 2, Appli
3	91	95.8	1375	4 US-09-210-361-4	Sequence 4, Appli
4	65	68.4	1430	3 US-09-008-172-2	Sequence 2, Appli
5	65	68.4	1430	4 US-09-210-361-6	Sequence 6, Appli
6	61	64.2	1577	2 US-08-793-824-2	Sequence 2, Appli
7	42	44.2	2627	2 US-08-751-189-3	Sequence 3, Appli
8	42	44.2	2627	2 US-09-060-836-3	Sequence 3, Appli
9	42	44.2	2627	4 US-09-184-445-3	Sequence 3, Appli
10	41.5	43.7	713	2 US-08-849-212-4	Sequence 4, Appli
11	40	42.1	431	1 US-07-783-705A-2	Sequence 2, Appli
12	38.5	40.5	237	2 US-08-578-709-11	Sequence 11, Appli
13	38.5	40.5	345	3 US-09-231-529-1	Sequence 1, Appli
14	38.5	40.5	345	4 US-08-977-816-1	Sequence 1, Appli
15	38.5	40.5	500	2 US-08-578-709-15	Sequence 15, Appli
16	38	40.0	184	1 US-08-353-550-2	Sequence 2, Appli
17	38	40.0	184	2 US-08-551-687-2	Sequence 2, Appli
18	38	40.0	428	1 US-08-353-550-1	Sequence 1, Appli
19	38	40.0	428	2 US-08-551-687-1	Sequence 1, Appli
20	38	40.0	431	3 US-08-807-342B-5	Sequence 5, Appli
21	38	40.0	469	1 US-08-353-550-6	Sequence 6, Appli
22	38	40.0	469	2 US-08-551-687-6	Sequence 6, Appli
23	38	40.0	639	2 US-08-557-309B-37	Sequence 37, Appli
24	38	40.0	639	3 US-08-834-306-37	Sequence 37, Appli
25	38	40.0	639	4 US-08-993-674A-37	Sequence 37, Appli
26	37	38.9	334	2 US-08-566-096A-6	Sequence 6, Appli
27	37	38.9	334	2 US-08-668-650B-6	Sequence 6, Appli

28	37	38.9	334	5 PCT-US95-15646-6	Sequence 6, Appli
29	37	38.9	445	2 US-08-630-118A-2	Sequence 2, Appli
30	37	38.9	445	2 US-08-630-118A-4	Sequence 4, Appli
31	37	38.9	445	2 US-08-838-399-2	Sequence 2, Appli
32	37	38.9	445	2 US-08-838-399-4	Sequence 4, Appli
33	37	38.9	445	2 US-09-003-199-23	Sequence 23, Appli
34	37	38.9	445	4 US-09-235-839-2	Sequence 4, Appli
35	37	38.9	445	4 US-09-235-839-4	Sequence 4, Appli
36	37	38.9	456	1 US-08-349-025-2	Sequence 2, Appli
37	37	38.9	456	2 US-08-566-096A-2	Sequence 2, Appli
38	37	38.9	456	2 US-08-668-650B-2	Sequence 2, Appli
39	37	38.9	456	2 US-08-668-650B-14	Sequence 14, Appli
40	37	38.9	456	5 PCT-US95-15646-2	Sequence 2, Appli
41	36.5	38.4	613	1 US-08-405-615-1	Sequence 1, Appli
42	36.5	38.4	613	2 US-08-461-234-1	Sequence 1, Appli
43	36.5	38.4	613	2 US-08-463-480-1	Sequence 1, Appli
44	36.5	38.4	614	1 US-08-225-224-1	Sequence 1, Appli
45	36.5	38.4	614	3 US-08-722-258-1	Sequence 1, Appli

#### ALIGNMENTS

```
RESULT 1
US-09-007-999-2
: Sequence 2, Application US/09007999
: Patent No. 6087559
: GENERAL INFORMATION:
: APPLICANT: Nichols, Scott E.
: TITLE OF INVENTION: Substitutes for Modified Starch and
: FILE REFERENCE: 0356D
: CURRENT APPLICATION NUMBER: US/09/007, 999
: EARLIER FILING DATE: 1998-01-16
: EARLIER APPLICATION NUMBER: 08/478, 704
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 1475
: TYPE: PRT
: ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match      100.0%; Score 95; DB 3; Length 1475;
Best Local Similarity 100.0%; Pred. No. 9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVODLIA 19
Db 552 VPSYSFIRAHSEVODLIA 570

RESULT 2
US-09-210-361-2
: Sequence 2, Application US/09210361
: Patent No. 6284479
: GENERAL INFORMATION:
: APPLICANT: Nichols, Scott E.
: TITLE OF INVENTION: Substitutes for Modified Starches and
: FILE REFERENCE: 0357CR
: CURRENT APPLICATION NUMBER: US/09/210, 361
: EARLIER FILING DATE: 1998-12-11
: EARLIER APPLICATION NUMBER: 09/007, 999
: EARLIER FILING DATE: 1998-01-16
: EARLIER APPLICATION NUMBER: 08/478, 704
: EARLIER FILING DATE: 1995-06-07
: EARLIER APPLICATION NUMBER: 09/009, 620
: EARLIER FILING DATE: 1998-01-20
: EARLIER APPLICATION NUMBER: 08/485, 243
: EARLIER FILING DATE: 1995-06-07
```

```
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2
```

```
Query Match          100.0%; Score 95; DB 4; Length 1475;
Best Local Similarity 100.0%; Pred. No. 9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
1 VPSYSFIRAHDESEVODLIA 19
|||||
552 VPSYSFIRAHDESEVODLIA 570
```

```
RESULT 3
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4
```

```
Query Match          95.8%; Score 91; DB 4; Length 1375;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 VPSYSFIRAHDESEVODLI 18
|||||
DB 578 VPSYSFIRAHDESEVODLI 595
```

```
RESULT 4
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
```

```
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2
```

```
Query Match          68.4%; Score 65; DB 3; Length 1430;
Best Local Similarity 76.5%; Pred. No. 0.0026;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 3 SYSFIRAHDESEVODLIA 19
|||||
DB 576 NYIFIRAHDESEVOTVIA 592
```

```
RESULT 5
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6
```

```
Query Match          68.4%; Score 65; DB 4; Length 1430;
Best Local Similarity 76.5%; Pred. No. 0.0026;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 3 SYSFIRAHDESEVODLIA 19
|||||
DB 576 NYIFIRAHDESEVOTVIA 592
```

```
RESULT 6
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; INCREASE STARCH CONTENT
; NUMBER OF SEQUENCES: 2
```



US-09-184-445-3  
; Sequence 3, Application US/09184445  
; Patent No. 6174703  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, Lea A.  
; APPLICANT: Robinson, Murray O.  
; TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein  
; TITLE OF INVENTION: 1  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen, Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/184,445  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/751,189  
; FILING DATE: 15-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oleski, Nancy A.  
; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-433  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2627 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-09-184-445-3

Query Match 44.2%; Score 42; DB 4; Length 2627;  
Best Local Similarity 47.1%; Pred. No. 85;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

1 VPSYSTRADSEVQDL 17  
:||||| : |||||  
205 MPYSYSLGEEVEEDL 221

RESULT 10  
US-08-849-212-4  
; Sequence 4, Application US/08849212  
; Patent No. 5827698  
; GENERAL INFORMATION:  
; APPLICANT: KIKUCHI, YOSHIMI  
; APPLICANT: SUZUKI, TOMOKO  
; APPLICANT: KOTIMA, HIROUKI  
; TITLE OF INVENTION: NOVEL LYSINE DECARBOXYLASE GENE AND  
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/849,212  
; FILING DATE: 09-JUN-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6/306386  
; FILING DATE: 09-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 10-856-0PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 713 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-849-212-4

Query Match 43.7%; Score 41.5; DB 2; Length 713;  
Best Local Similarity 45.0%; Pred. No. 22;  
Matches 9; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

QY 1 VPSYSTRADSEVQDL 17  
Db 77 LPLVAFINTHSTWDVSQDM 96

RESULT 11  
US-07-783-705A-2  
; Sequence 2, Application US/07783705A  
; Patent No. 5429939  
; GENERAL INFORMATION:  
; APPLICANT: Misawa, No. 5429939ihiko  
; APPLICANT: Kobayashi, Kazuo  
; APPLICANT: Nakamura, Katsunori  
; APPLICANT: Yamano, Shigeyuki  
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE  
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOID  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ladas & Parry  
; STREET: 26 West 61 Street  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10023  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720kb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: N/A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/783,705A  
; FILING DATE: 19911023  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 1-103078  
; FILING DATE: 21-APR-1989  
; APPLICATION NUMBER: JP 2-53225  
; FILING DATE: 05-MAR-1990  
; APPLICATION NUMBER: US 07/519,011  
; FILING DATE: 19-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Schwadron, Janet I.  
; REGISTRATION NUMBER: 33,778  
; TELECOMMUNICATION INFORMATION:



; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/977,816  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0429 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 345 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: KIDNNOT25  
; CLONE: 353694  
; US-08-977-816-1

Query Match 40.5%; Score 38.5; DB 4; Length 345;  
Best Local Similarity 40.0%; Pred. No. 32;  
Matches 8; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 1 VPSYSFI--RAHSEVODL 17  
||| : ||| : ||| :  
DB 111 VPSLEKVLKFAHEGIEDL 130

QY 15  
08-578-709-15  
; Sequence 15, Application US/08578709  
; Patent No. 5814509  
; GENERAL INFORMATION:  
; APPLICANT: TANABE, Tadashi  
; TITLE OF INVENTION: PROSTACYCLIN SYNTHASE DERIVED FROM HUMAN  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
; STREET: 2100 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/578,709  
; FILING DATE: 28-DEC-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP95/00838

; FILING DATE: 27-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 114316/1994  
; FILING DATE: 28-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gubinsky, Louis  
; REGISTRATION NUMBER: 24,835  
; REFERENCE/DOCKET NUMBER: 040439  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)293-7060  
; TELEFAX: (202)293-7860  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 500 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-578-709-15

Query Match 40.5%; Score 38.5; DB 2; Length 500;  
Best Local Similarity 29.0%; Pred. No. 50;  
Matches 9; Conservative 4; Mismatches 3; Indels 15; Gaps 1;

QY 3 SYSFI-----RAHSEVODLI 18  
||| : ||| : ||| :  
DB 168 SYSFLRAGYLTLTGIEALPRHESQADRV 198

Search completed: March 27, 2002, 13:59:27  
Job time: 580 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:00 ; Search time 53.4 Seconds  
(without alignments)  
13.046 Million cell updates/sec

Title: US-09-290-049a-2  
Perfect score: 95  
Sequence: 1 VPSYSTRAHDSVQDLIA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Number of hits satisfying chosen parameters: 100059

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	95.8	1375	1	GTFC_STRMU
2	91	95.8	1476	1	GTFC_STRMU
3	86	90.5	1597	1	GTFC_STRMU
4	84	88.4	1592	1	GTFC_STRMU
5	73	76.8	1365	1	GTFC_STRMU
6	65	68.4	1462	1	GTFC_STRMU
7	43	45.3	214	1	COMA_BACSU
8	42	44.2	644	1	RNB_ECOLI
9	42	44.2	659	1	RNB_HAEIN
10	42	44.2	932	1	HIRL_SCHPO
11	41.5	43.7	713	1	DLIZ_ECOLI
12	41	43.2	344	1	MIZD_BACSU
13	40	42.1	205	1	ADEN_ADEG8
14	40	42.1	431	1	CRTX_PANAN
15	40	42.1	548	1	LIP2_CANRU
16	40	42.1	549	1	LIP3_CANRU
17	40	42.1	549	1	LIP4_CANRU
18	40	42.1	572	1	PTLB_STANU
19	40	42.1	6359	1	BACC_BACLI
20	39.5	41.6	192	1	CUPP_DROPS
21	39	41.1	403	1	COAT_BOOLV
22	39	41.1	509	1	CAAS_DROME
23	39	41.1	585	1	ASNS_MAIZE
24	39	41.1	715	1	DOAL_YEAST
25	39	41.1	737	1	SKNI_CANAL
26	39	41.1	946	1	YB16_YEAST
27	39	41.1	1389	1	PYAX_RAT
28	39	41.1	2090	1	N214_HUMAN
29	39	41.1	2116	1	MYS2_DICDI
30	38.5	40.5	454	1	NFM_PLG
31	38.5	40.5	494	1	AINX_HUMAN
32	38.5	40.5	500	1	PTGI_HUMAN
33	38.5	40.5	504	1	AINX_MOUSE

## ALIGNMENTS

RESULT ID	1	GTFC_STRMU	STANDARD:	PRT: 1375 AA.	
AC	P13470: P05427:				
DT	01-NOV-1988 (rel. 09, Created)				P23565 rattus norv
DT	01-JAN-1990 (rel. 13, Last sequence update)				O77788 bos taurus
DT	15-DEC-1998 (rel. 37, Last annotation update)				P12839 rattus norv
DE	GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)				P07197 homo sapien
DE	(DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).				O57601 methanococ
GN	GTFC.				P37270 thermus aqu
OS	Streptococcus mutans.				O58890 methanococ
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				O83207 murine cyto
OX	Streptococcus.				P98077 homo sapien
OX	NCBI_TaxId=1309;				P11714 mus musculu
RP	SEQUENCE FROM N.A.				O92345 schizosach
RC	STRAIN=GS-5;				P46509 mycobacteri
RX	MEDLINE=89137980; PubMed=2976010;				
RA	Ueda S., Shiroza T., Kuramitsu H.K.;				
RT	"Sequence analysis of the gtfc gene from Streptococcus mutans GS-5.";				
RL	Gene 69:101-109(1988).				
RN	[2]				
RP	SEQUENCE OF 1-349 FROM N.A.				
RC	STRAIN=GS-5;				
RX	MEDLINE=87308013; PubMed=3040685;				
RA	Shiroza T., Ueda S., Kuramitsu H.K.;				
RT	"Sequence analysis of the gtfb gene from Streptococcus mutans.";				
RL	J. Bacteriol. 163:4263-4270(1987)				
CC	- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.				
CC	- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) - D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).				
CC	- SUBCELLULAR LOCATION: SECRETED.				
CC	- DISEASE: DENTAL CARIES.				
CC	- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOS). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.				
CC	- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S. MUTANS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a> or send an email to <a href="mailto:license@isb.sib.ch">license@isb.sib.ch</a> ).				
CC	-----				
DR	EMBL: M22054; AA88592.1; -				
DR	EMBL: M17361; AA88589.1; -				
DR	PIR: J070345; J070345.				
DR	PIR: C31335; C31335.				
DR	InterPro: IPR002479; CW_binding.				

cc between

SO	SEQUENCE	1476	AA:	165685	MW:	3479862807694D98	CRC64:
FM	CONFLICT	1310		1310	H -> L (IN REF. 1).		

between the Swiss Institute of Bioinformatics and the EMBL outstation -

Query Match 95.8%; Score 91; DB 1; Length 1476;  
 Best Local Similarity 100.0%; Pred. No. 2,9e-07;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVODLI 18  
 |||||||  
 DB 552 VPSYSFIRAHSEVODLI 569

RESULT 3  
 GTF1\_STRDO STANDARD; PRT: 1597 AA.

DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).

GTF1.  
 Streptococcus downei (Streptococcus sobrinus).  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.

OX NCBI\_TaxID=1317;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MEF28;

RA MEDLINE=87308014; PubMed=3040686;

RT Ferretti J.J., Gilpin M.L., Russell R.R.B.;

RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus

sobrinus MEF28";

RL J. Bacteriol. 169:4271-4278(1987).

CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT

TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE

OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE

AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.

CC -I- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =

D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).

CC -I- SUBCELLULAR LOCATION: SECRETED.

CC -I- DISEASE: DENTAL CARIES.

CC -I- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA

1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES

WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH

FORMS OF GLUCANS.

CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-

BINDING PROTEIN FROM S. MUTANS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M17391; AAC63063.1; -

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR003318; Glyco\_hydro\_70.

DR Pfam: PF01473; CW\_binding\_1; 19.

DR Pfam: PF02324; Glyco\_hydro\_70; 1.

KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

KW SIGNAL 1 38 POTENTIAL.

FT CHAIN 39 1597 GLUCOSYLTRANSFERASE-I.

FT DOMAIN 39 1050 CATALYTIC (APPROXIMATE).

FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).

FT DOMAIN 1099 1597 1-25 A, 2 B, AND 5 AC REPEATS.

FT REPEAT 1099 1132 A REPEAT.

FT REPEAT 1163 1213 AC REPEAT.

FT REPEAT 1227 1277 AC REPEAT.

FT REPEAT 1292 1342 AC REPEAT.

FT REPEAT 1352 1399 B REPEAT.

FT REPEAT 1406 1455 AC REPEAT.

FT REPEAT 1465 1512 B REPEAT.  
 FT REPEAT 1519 1568 AC REPEAT.  
 FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).  
 SQ SEQUENCE 1597 AA; 177080 MW; B9E86A20868798E CRC64;

Query Match 90.5%; Score 86; DB 1; Length 1597;  
 Best Local Similarity 94.4%; Pred. No. 2.2e-06;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVODLI 18  
 |||||||  
 DB 554 VPSYSFIRAHSEVODLI 571

RESULT 4  
 GTF2\_STRDO STANDARD; PRT: 1592 AA.

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)

DE (SUCROSE 6-GLUCOSYLTRANSFERASE).

OS Streptococcus downei (Streptococcus sobrinus).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1317;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=6715;

RA MEDLINE=91123227; PubMed=1704006;

RT Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,

RA Kawada H.;

RT "peptide sequences for sucrose splitting and glucan binding within

Streptococcus sobrinus glucosyltransferase (water-insoluble glucan

synthetase)";

RL J. Bacteriol. 173:989-996(1991).

CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT

TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE

OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE

AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.

CC -I- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =

D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).

CC -I- SUBCELLULAR LOCATION: SECRETED.

CC -I- DISEASE: DENTAL CARIES.

CC -I- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA

1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES

WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH

FORMS OF GLUCANS.

CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-

BINDING PROTEIN FROM S. MUTANS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: D90213; BA014241.1; -

DR PIR: A38175; A38175.

DR HSSP: P00695; 2HEE.

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR003318; Glyco\_hydro\_70.

DR Pfam: PF01473; CW\_binding\_1; 16.

DR Pfam: PF02324; Glyco\_hydro\_70; 1.

KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

KW SIGNAL 1 38 POTENTIAL.

FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-I.

FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).

FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).

FT DOMAIN 1093 1592 6.5 X TANDEM REPEATS.  
FT REPEAT 1093 1142 1.  
FT REPEAT 1158 1207 2.  
FT REPEAT 1222 1272 3.  
FT REPEAT 1287 1337 4.  
FT REPEAT 1402 1451 5.  
FT REPEAT 1514 1563 6.  
FT REPEAT 1577 1592 7 (INCOMPLETE).  
SO SEQUENCE 1592 AA; 176167 MW; BC0A66D079351BECF CRC64;

Query Match 88.4%; Score 84; DB 1; Length 1592;  
Best Local Similarity 88.9%; Pred. No. 4,8e-06;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPYSFIRAHDSFVODLI 18  
DB 548 VPYSFIRAHDSFVODII 565

RESULT 5  
GTF5\_STRDO STANDARD: PRT: 1365 AA.  
ID GTF5\_STRDO  
AC P29336;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
GN GTF5.  
OS Streptococcus downei (Streptococcus sobrinus).  
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MFE28;  
RA MEDLINE=90316665; PubMed=2142479;  
RX Gilmore K.S., Russell R.R., Ferretti J.J.;  
RT "Analysis of the Streptococcus downei gtf5 gene, which specifies a  
glucosyltransferase that synthesizes soluble glucans."  
RL Infect. Immun. 58:2452-2458(1990).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
PRIMER GLUCAN UNLIKE GTF-I.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA  
1,6-GLUCOSE).  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M30943; AAA26898.1; -  
CC PIR: A41483; A41483.  
CC InterPro: IPR002479; CM\_binding.  
CC InterPro: IPR003318; Glyco\_hydro\_70.  
CC Pfam: PF01473; CM\_binding\_1; 10.  
CC Pfam: PF02324; Glyco\_hydro\_70; 1.  
CC Transferrase: Glycosyltransferase; Signal; Repeat; Dental caries.  
KW SIGNAL 1  
FT CHAIN 37 1365 OR 37 (POTENTIAL).  
FT SIGNAL 36 GLUCOSYLTRANSFERASE-S.

FT DOMAIN 37 1050 CATALYTIC (APPROXIMATE).  
FT REPEAT 1083 1365 GLUCAN-BINDING (APPROXIMATE).  
FT DOMAIN 1083 1365 4.5 X TANDEM REPEATS.  
FT REPEAT 1083 1131 1.  
FT REPEAT 1150 1199 2.  
FT REPEAT 1225 1274 3.  
FT REPEAT 1289 1339 4.  
FT REPEAT 1353 1365 5 (INCOMPLETE).  
SO SEQUENCE 1365 AA; 151590 MW; 167296B5A2E8C476 CRC64;

Query Match 76.8%; Score 73; DB 1; Length 1365;  
Best Local Similarity 78.9%; Pred. No. 0.00029;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 VPYSFIRAHDSFVODLIA 19  
DB 537 VPNYFIRAHDSFVOTRIA 555

RESULT 6  
GTFD\_STRMU STANDARD: PRT: 1462 AA.  
ID GTFD\_STRMU  
AC P49331; O69383; O69386; O69389; O69392; O69398;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
GN GTFD.  
OS Streptococcus mutans.  
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RA MEDLINE=91100958; PubMed=2148600;  
RX Honda O., Kato C., Kuramitsu H.K.;  
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
the glucosyltransferase-S enzyme."  
RL J. Gen. Microbiol. 136:2099-2105(1990).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC STRAIN-MT429, MT4245, MT4251, MT4467, AND MT8148;  
CC MEDLINE=98231643; PubMed=9570124;  
CC Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
CC Kimura S., Hamada S.;  
CC "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans."  
CC FEWS Microbiol. Lett. 161:331-336(1998).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
FORMS OF GLUCANS.  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

```

CC EMBL: M29296: AAA26895.1; -
DR EMBL: D88653: BAA26103.1; -
DR EMBL: D88656: BAA26107.1; -
DR EMBL: D88659: BAA26111.1; -
DR EMBL: D88662: BAA26115.1; -
DR EMBL: D89979: BAA26121.1; -
DR InterPro: IPR002479: CW-binding.
DR InterPro: IPR003318: Glyco_hydro_70.
DR Pfam: PF01473: CW_binding_1; 11.
DR Pfam: PF03324: Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1
FT CHAIN ? 1462 GLUCOSYLTRANSFERASE-S.
FT DOMAIN 1232 1423 3 X 63 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 1232 1295 1.
FT REPEAT 1296 1359 2.
FT REPEAT 1360 1423 3.
FT VARIANT 58 58 K -> E (IN STRAIN MT4467).
FT VARIANT 68 68 A -> S (IN STRAIN MT4239 AND MT4245).
FT VARIANT 81 81 A -> T (IN STRAIN MT4251 AND MT8148).
FT VARIANT 113 113 T -> I (IN STRAIN MT4239 AND MT4245).
FT VARIANT 122 122 A -> V (IN STRAIN MT4239, MT4245 AND MT8148).
FT VARIANT 132 132 S -> A (IN STRAIN MT4239, MT4245, MT4251 AND MT8148).
FT VARIANT 135 135 A -> V (IN STRAIN MT4245).
FT VARIANT 202 202 V -> L (IN STRAIN MT4239).
FT VARIANT 255 255 D -> N (IN STRAIN MT8148).
FT VARIANT 275 275 E -> D (IN STRAIN MT4239, MT4245 AND MT4251).
FT VARIANT 288 288 D -> N (IN STRAIN MT4239, MT4245 AND MT4251).
FT VARIANT 301 301 Q -> H (IN STRAIN MT4245).
FT VARIANT 313 313 D -> N (IN STRAIN MT4239 AND MT4251).
FT VARIANT 317 317 E -> F (IN STRAIN MT4239).
FT VARIANT 328 328 V -> K (IN STRAIN MT4239).
FT VARIANT 350 350 F -> L (IN STRAIN MT4239, MT4251 AND MT4467).
FT VARIANT 628 628 KKRIQ -> EKEVTL (IN STRAIN MT4251).
FT VARIANT 633 633 A -> S (IN STRAIN MT4239).
FT VARIANT 688 688 TDGSGEA -> ADKNGDS (IN STRAIN MT4251).
FT VARIANT 726 726 TDGGS -> ADKGN (IN STRAIN MT4239 AND MT4245).
FT VARIANT 964 964 D -> Y (IN STRAIN MT4251).
FT VARIANT 1019 1019 E -> K (IN STRAIN MT4245 AND MT4251).
FT VARIANT 1059 1059 LG -> IR (IN STRAIN MT4251).
FT VARIANT 1060 1060 G -> R (IN STRAIN MT4245).
FT VARIANT 1080 1080 G -> R (IN STRAIN MT4239).
FT VARIANT 1142 1142 O -> H (IN STRAIN MT4239, MT4245, MT4251, MT4467 AND MT8148).
FT VARIANT 1198 1198 S -> N (IN STRAIN MT4239).
FT VARIANT 1220 1220 Y -> C (IN STRAIN MT4251 AND MT4467).
FT VARIANT 1280 1280 F -> L (IN STRAIN MT4467).
FT VARIANT 1282 1282 O -> P (IN STRAIN MT4245).
FT VARIANT 1290 1290 K -> T (IN STRAIN MT4245).
FT VARIANT 1311 1311 N -> D (IN STRAIN MT4245).
FT VARIANT 1403 1403 D -> G (IN STRAIN MT4239, MT4245, MT4251 AND MT8148).
FT VARIANT 1425 1425 R -> G (IN STRAIN MT4239, MT4245, MT4251, MT4467 AND MT8148).
FT VARIANT 1449 1449 R -> K (IN STRAIN MT4467).
FT CONFLICT 1428 1462 RYDKNSGNMYNKVTLANGRLIGIDRWGIARY -> VYR (IN REF. 1).
SQ SEQUENCE 1462 AA; 163512 MW; 5C6541F0DCB0DF00 CRC64;

```

```

Query Match 68.4%; Score 65; DB 1; Length 1462;
Best Local Similarity 76.5%; Pred. No. 0.007;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
3 SYSFIRAHSEVQTVIA 19
: |||||

```

```

DB 576 NYFIRAHSEVQTVIA 592
RESULT 7
COMA_BACSU STANDARD; PRT; 214 AA.
ID COMA_BACSU
AC PI4204;
DT 01-JAN-1990 (Rel. 13, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE COMPETENCE PROTEIN A.
GN COMA OR COMAA OR COMA1.
OS Bacillus subtilis.
OC Bacillus; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90008771; PubMed=2507523;
RA Weinrauch Y., Guillen N., Dubnau D.;
RT "Sequence and transcription mapping of Bacillus subtilis competence
RT genes comb and coma, one of which is related to a family of bacterial
RT regulatory determinants."
RL J. Bacteriol. 171:5362-5375(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Oudega B., Koningssteijn G., van Zoest A.;
RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: REQUIRED FOR THE DEVELOPMENT OF COMPETENCE IN B.
CC SUBTILIS. FOR THE EXPRESSION OF LATE-EXPRESSING COMPETENCE
CC GENES. AND FOR THE EXPRESSION OF THE GROWTH STAGE-REGULATED
CC MOLECULE SURFACTIN. MAY PLAY A REGULATORY ROLE DURING THE
CC DEVELOPMENT OF COMPETENCE, AND AN ANALOGOUS ROLE AS UVRC-ORF2
CC (TRANSDUCING ENVIRONMENTAL INFORMATION TO THE DNA REPAIR SYSTEM).
CC -I- PTM: PHOSPHORYLATED BY COMP.
CC -I- SIMILARITY: TO OTHER BACTERIAL REGULATORY PROTEINS INVOLVED IN
CC SIGNAL TRANSDUCTION.
CC -I- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M2856: AAA22320.1; -
DR EMBL: Z93932: CAB07904.1; -
DR EMBL: Z99120: CAB15156.1; -
DR PIR: A33591: RGSBCA.
DR Subtilist: BG10381: COMA.
DR InterPro: IPR001789; Response_reg.
DR Pfam: PF00196; Gere; 1.
DR SMART: SMO0472; response_reg; 1.
DR SMART: SMO0448; REC; 1.
DR PROSITE: PS00622; HTH_LUXR_FAMILY; 1.
KW Sensory transduction; Phosphorylation; Transcription regulation;
KW DNA-binding; Activator; Complete proteome.
FT DOMAIN 1 101 RECEIVER DOMAIN.
FT MOD_RES 55 55 PHOSPHORYLATION (BY SIMILARITY).
FT DNA_BIND 171 190 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 214 AA; 24128 MW; 52D01A9740759072 CRC64;

```

```

Query Match 45.3%; Score 43; DB 1; Length 214;
Best Local Similarity 52.9%; Pred. No. 4.4;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

```

```

OY      2 PSYFIRAHSEVODLI 18
DB      36 PSYFIRAHSEVODLI 52

RESULT 8
RNB_ECOLI STANDARD: PRT: 644 AA.
AC P30850: P78280:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXORIBONUCLEASE II (EC 3.1.13.1) (RIBONUCLEASE II) (RNASE II).
GN RNB OR B1286.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
   Escherichia.
NCBI_TaxID=562;
(1)
SEQUENCE FROM N.A.
RC STRAIN-K12:
RA MEDLINE=93268098; PubMed=8497196;
RA Zilhao R., Cameio L., Arraiano C.M.;
RT "DNA sequencing and expression of the gene rnb encoding Escherichia
RT coli ribonuclease II."
RL Mol. Microbiol. 8:43-51(1993).
RN [2]
RP REVISIONS.
RA Zilhao R.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12:
RA MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasei H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Samedel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horluchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
RN [5]
RP FUNCTION: INVOLVED IN MRNA DEGRADATION. HYDROLYZES SINGLE-STRANDED
RP POLYRIBONUCLEOTIDES PROCESSIVELY IN THE 3' TO 5' DIRECTION.
RC CATALYTIC ACTIVITY: EXONUCLEOTIC CLEAVAGE IN THE 3' TO 5'
CC DIRECTION TO YIELD 5'-PHOSPHOMONONUCLEOTIDES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.
CC -1- SIMILARITY: CONTAINS 1 'SI MOTIF' DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 'SI MOTIF' DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
EMBL: X67913; CAA8112.1;
EMBL: AEO00226; AAC74368.1;
EMBL: D90767; BAA14848.1;
EMBL: D90766; BAA14840.1;

```

```

DR PIR: S28506; S28506.
DR PIR: S32940; S32940.
DR Ecogene; Bg11620; rnb.
DR InterPro; IPR002059; Cold_shock.
DR InterPro; IPR001900; Ribonuclease_II.
DR InterPro; IPR003029; S1.
DR Pfam; PF00773; RNB; 1.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00357; CSP; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PS01175; RIBONUCLEASE_II; 1.
DR Hydrolase; Exonuclease; Nuclease; RNA-binding; Complete proteome.
KW CONFLICT 384 384 I -> N (IN REF. 1).
FT CONFLICT 399 399 C -> R (IN REF. 1).
FT CONFLICT 513 513 A -> R (IN REF. 1).
SQ SEQUENCE 644 AA; 72490 MW; 36B1671ZCDF14394 CRC64;

Query Match 44.2%; Score 42; DB 1; Length 644;
Best Local Similarity 44.4%; Pred. No. 22;
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY      2 PSYFIRAHSEVODLI 19
DB      362 PSYFIRAHSEVODLI 379

RESULT 9
RNB_HAEIN STANDARD: PRT: 659 AA.
AC P44440;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXORIBONUCLEASE II (EC 3.1.13.1) (RIBONUCLEASE II) (RNASE II).
GN RNB OR H11733.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
   Haemophilus.
NCBI_TaxID=727;
(1)
RP SEQUENCE FROM N.A.
RC STRAIN-RD / ATCC 51907;
RA MEDLINE=95350630; PubMed=7542800;
RA Fletschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Goeyne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrich T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fritschman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
RN [1]
RP FUNCTION: INVOLVED IN MRNA DEGRADATION. HYDROLYZES SINGLE-STRANDED
RP POLYRIBONUCLEOTIDES PROCESSIVELY IN THE 3' TO 5' DIRECTION
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: EXONUCLEOTIC CLEAVAGE IN THE 3' TO 5'
CC DIRECTION TO YIELD 5'-PHOSPHOMONONUCLEOTIDES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.
CC -1- SIMILARITY: CONTAINS 1 'SI MOTIF' DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 'SI MOTIF' DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```



RT mutations in essential genes by modulating Rho-dependent  
 RT transcription termination."  
 RL Mol. Microbiol. 29:859-869(1998).  
 CC -1- FUNCTION: LDC IS CONSTITUTIVELY BUT WEAKLY EXPRESSED UNDER VARIOUS  
 CC CONDITIONS. OPTIMUM ACTIVITY IS ACHIEVED BETWEEN PH 6.2 TO 8.0.  
 CC -1- CATALYTIC ACTIVITY: L-LYSINE - CADAVERINE + CO(2).  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- SUBUNIT: HOMODECAMER.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF ORNITHINE, LYSINE, AND ARGinine  
 CC DECARBOXYLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D87518; BAA21656.1; -  
 DR EMBL: D49445; BAA08426.1; -  
 DR EMBL: AE000128; AAC73297.1; -  
 DR EMBL: D83536; BAA77861.1; -  
 DR EMBL: U70214; AAB08615.1; -  
 DR EMBL: Z50870; CAA90749.1; -  
 DR Ecocore: E813219; ldcC.  
 DR InterPro: IPR000310; Orn\_Lys\_Arg\_decarboxylase\_1.  
 DR Pfam: PF01276; OKR\_DC\_1; 1.  
 DR PROSITE: PS00703; OKR\_DC\_1; 1.  
 KW Lyase; Decarboxylase; Pyridoxal phosphate; Multigene family;  
 KW Complete proteome.  
 FT BINDING 367 PYRIDOXAL PHOSPHATE.  
 FT CONFLICT 284 S -> T (IN REF. 2).  
 FT CONFLICT 314 N -> F (IN REF. 2).  
 FT CONFLICT 411 T -> S (IN REF. 2).  
 FT CONFLICT 413 AA -> R (IN REF. 2).  
 FT CONFLICT 498 AA -> I (IN REF. 2).  
 FT CONFLICT 673 L -> I (IN REF. 2).  
 SO SEQUENCE 713 AA; 80590 MW; 4532C6069744ABDF CRC64;

Query Match 43.7%; Score 41.5; DB 1; Length 713;  
 Best Local Similarity 45.0%; Pred. No. 29;  
 Matches 9; Conservative 3; Mismatches 5; Indels 3; Gaps 1;  
 QY 1 VPSYSFIRAH---DSEVQDL 17  
 77 LPLVAFINTHSTMDVSVDQM 96

RESULT 12  
 MID2\_BACSU STANDARD: PRT: 344 AA.  
 AC P26935;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MYO-INOSITOL 2-DEHYDROGENASE (EC 1.1.1.18).  
 GN IDH OR IOLG OR E83G.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=60015;  
 RX MEDLINE=92104493; PubMed=1761221;  
 RA Fujita Y., Shindo K., Miwa Y., Yoshida K.;  
 RT "Bacillus subtilis inositol dehydrogenase-encoding gene (idh):  
 RT sequence and expression in Escherichia coli."  
 RL Gene 108:121-125(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=168 / BGSC1A1;  
 RX MEDLINE=95039891; PubMed=7952181;  
 RA Yoshida K., Sano H., Miwa Y., Ogasawara N., Fujita Y.;  
 RT "Cloning and nucleotide sequencing of a 15 kb region of the Bacillus  
 RT subtilis genome containing the iol operon."  
 RL Microbiology 140:2289-2298(1994).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=79239346; PubMed=112095;  
 RA Ramaley R., Fujita Y., Freese E.;  
 RT "Purification and properties of Bacillus subtilis inositol  
 RT dehydrogenase."  
 RL J. Biol. Chem. 254:7684-7690(1979).  
 CC -1- CATALYTIC ACTIVITY: MYO-INOSITOL + NAD(+) -> 2,4,6/3,5-  
 CC PENTAHYDROXYCYCLOHEXANONE + NADH.  
 CC -1- PATHWAY: FIRST STEP OF MYO-INOSITOL CATABOLISM.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- INDUCTION: INOSITOL, SUBJECTED TO CATABOLITE REPRESSION.  
 CC -1- SIMILARITY: BELONGS TO THE GFO/IDH/MOCA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M76431; AAA22543.1; -  
 DR EMBL: D14399; BAA03296.1; -  
 DR EMBL: Z99124; CAB16006.1; -  
 DR PIR: JH0511; JH0511.  
 DR Subtilist: BG10669; idh.  
 DR InterPro: IPR000683; GFO\_IDH\_MOCA.  
 DR Pfam: PF01408; GFO\_IDH\_MOCA; 1.  
 KW Oxidoreductase; NAD; Complete proteome.  
 SO SEQUENCE 344 AA; 38351 MW; 2FCE908DAE2C332P CRC64;

Query Match 43.2%; Score 41; DB 1; Length 344;  
 Best Local Similarity 53.8%; Pred. No. 16;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 6 FIRAHSEVQDLI 18  
 Db 276 EVAADVEIQDFI 288

RESULT 13  
 ADEN\_ADEG8 STANDARD: PRT: 205 AA.  
 AC Q90M72;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ADENAIN (EC 3.4.22.39) (ENDOPROTEASE) (LATE L3 23 KDA PROTEIN).  
 OS Avian adenovirus type 8 (strain AFCC A-2A) (Fowl adenovirus 8).  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadnavirus.  
 OX NCBI\_TaxID=66295;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ojic D., Nagy E.;  
 RT "The DNA sequence of fowl adenovirus 8."  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIOL PROTEASE CLEAVING SPECIFIC GLY-ALA PEPTIDES IN A  
 CC NUMBER OF VIRAL PRECURSOR PROTEINS (TERMINAL, IIA, VI, VII, VIII,  
 CC IIA AND MO) WHICH ARE REQUIRED FOR VIRION MATURATION. ALSO CLEAVES  
 CC HOST CELLS CYTOSKELETAL KERATINS K7 AND K18.  
 CC -1- CATALYTIC ACTIVITY: CLEAVES PROTEINS OF THE ADENOVIRUS AND ITS  
 CC HOST CELL AT TWO CONSENSUS SITES: -YAA-XAA-GLY-1-XAA-AND -YAA-  
 CC XAA-GLY-XAA-1-GLY-(IN WHICH YAA IS MET, ILE OR LEU, AND XAA IS ANY  
 CC AMINO ACID).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C5.



```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC / between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF083975; MAD50345.2; -
DR HSSP; P03252; LAMP.
DR MEROPS; C05.001; -
DR InterPro; IPR000855; Peptidase_C5.
DR Pfam; PF00770; Peptidase_C5; 1.
DR PRINTS; PR00703; ADYENDOPHASE.
DR ProDom; PD003705; Peptidase_C5; 1.
DR HydroLase; Thiol protease; Late protein.
FW ACT_SITE 55 55 BY SIMILARITY.
FT ACT_SITE 72 72 BY SIMILARITY.
FT ACT_SITE 122 122 BY SIMILARITY.
SEQUENCE 205 AA; 23701 MW; 36F0700CDFB85F62 CRC64;

Query Match 42.1%; Score 40; DB 1; Length 205;
Best Local Similarity 63.6%; Pred. No. 13;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 5 SFIRAHSEVQ 15
DB 183 SFFRAHESELK 193

RESULT 14
CRTX_PANAN STANDARD; PRT; 431 AA.
AC P21686;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ZEXANTHIN GLUCOSYL TRANSFERASE (EC 2.4.1.-).
GN CRTX OR UGT101.
OS Pantosa ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantosa.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=20D3;
RX MEDLINE=91072214; PubMed=2254247;
Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,
Nakamura K., Harashina K.,
"Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway
by functional analysis of gene products expressed in Escherichia
coli".
RT
FT
RL J. Bacteriol. 172:6704-6712(1990).
CC -I- FUNCTION: CATALYZES THE GLYCOSYLATION REACTION WHICH CONVERTS
CC ZEXANTHIN TO ZEXANTHIN-BETA-DIGLUCOSIDE.
CC -I- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -I- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; D90087; BAA14125.1; -
DR PIR; B37802; B37802.
DR InterPro; IPR002213; UDPGT.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Carotenoid biosynthesis.

```

```

SQ SEQUENCE 431 AA; 47241 MW; 9EC27A798608BC9D CRC64;

Query Match 42.1%; Score 40; DB 1; Length 431;
Best Local Similarity 44.4%; Pred. No. 30;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 PSYFRAHSEVODLIA 19
DB 10 PSYSHKALQNLAEELVA 27

RESULT 15
LIP2_CANRU STANDARD; PRT; 548 AA.
ID LIP2_CANRU
AC P32946;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE LIPASE 2 PRECURSOR (EC 3.1.1.3).
GN LIP2.
OS Candida rugosa (Yeast) (Candida cylindracea).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5481;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14830;
RX MEDLINE=92305068; PubMed=1610906;
RA Longhi S., Fusetti F., Grandori R., Lotli M., Vanoni M.,
RA Albergina L.;
RT "Cloning and nucleotide sequences of two lipase genes from Candida
RT cylindracea."
RT Biochim. Biophys. Acta 1131:227-232(1992).
RL [2]
RN
RM REVIEW.
RX MEDLINE=98451816; PubMed=9778794;
RA Benjamin S., Pandey A.;
RT "Candida rugosa lipases: molecular biology and versatility in
RT biotechnology."
RL Yeast 14:1069-1087(1998).
CC -I- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +
CC A FATTY ACID ANION.
CC -I- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLLESTERASE/LIPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X64704; CAA45958.1; -
DR PIR; S32615; S32615.
DR HSSP; P32947; LCLE.
DR InterPro; IPR002018; Carboxylesterase_B.
DR InterPro; IPR000379; Est_lip_thioestl_actsite.
DR Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW HydroLase; Lipid degradation; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 14
FT CHAIN 15 548 LIPASE 2.
FT ACT_SITE 223 223 BY SIMILARITY.
FT ACT_SITE 355 355 BY SIMILARITY.
FT ACT_SITE 463 463 BY SIMILARITY.
FT DISULFD 74 111 BY SIMILARITY.
FT DISULFD 282 291 BY SIMILARITY.
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE 548 AA; 58864 MW; E0DBCFE2501E7614 CRC64;

```

Query Match 42.1%; Score 40; DB 1; Length 548;  
Best Local Similarity 53.3%; Pred. No. 39;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 SFRHDSVQDLIA 19  
||| | | | | | | | | |  
Db 379 SFRHDSVQDLIA 393

Search completed: March 27, 2002, 14:27:01  
Job time: 1643 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:01 ; Search time 188.53 Seconds  
(without alignments)  
14.741 Million cell updates/sec

Title: US-09-290-049a-2  
Perfect score: 95  
Sequence: 1 VPSYFIRAHSEVODLIA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Number of hits satisfying chosen parameters: 473505

Maximum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_17:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.podent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	95.8	1390	2	069385 streptococc
2	91	95.8	1455	2	069382 streptococc
3	91	95.8	1455	2	069388 streptococc
4	91	95.8	1455	2	069391 streptococc
5	91	95.8	1455	2	069397 streptococc
6	84	88.4	1590	2	055263 streptococc
7	84	88.4	1590	2	059983 streptococc
8	80	84.2	1477	2	091466 leuconostoc
9	80	84.2	1508	2	052224 leuconostoc
10	80	84.2	1508	2	09EZH5 leuconostoc
11	80	84.2	1527	2	09ZAR4 leuconostoc
12	79	83.2	1290	2	048756 leuconostoc
13	77	81.1	1016	2	091CJ7 leuconostoc
14	74	77.9	1512	2	09WXR5 streptococc
15	73	76.8	1338	2	09WXR4 streptococc
16	66	69.3	2057	2	09RE05 leuconostoc
17	65	68.4	1449	2	068542 streptococc
18	65	68.4	1449	2	055264 streptococc
19	65	68.4	1518	2	Q00600 streptococc

20	64	67.4	1575	2	09LCH3 streptococc
21	64	67.4	1577	2	054178 streptococc
22	64	67.4	1599	2	000599 streptococc
23	61	64.2	1577	2	055265 streptococc
24	50	52.6	93	2	09ZIX9 borrelia bu
25	46	48.4	66	2	09K5K6 bacillus mo
26	46	48.4	575	2	P90900 caenorhabdi
27	46	48.4	1032	10	09FIC9 arabidopsis
28	44	46.3	327	10	09C7I1 arabidopsis
29	44	46.3	371	8	09B8A8 arabidopsis
30	43	45.3	51	2	09RNP7 trichinella
31	43	45.3	64	2	09K5L0 bacillus su
32	43	45.3	1260	10	09LMD9 arabidopsis
33	42	44.2	247	10	09M8T9 enterobacte
34	42	44.2	431	2	047843 enterobacte
35	42	44.2	563	2	09JZL7 neisseria m
36	42	44.2	563	2	09JUP8 neisseria m
37	42	44.2	2627	4	099973 homo sapien
38	41.5	43.7	1882	3	09P898 podospira a
39	41	43.2	145	1	027766 methanobact
40	41	43.2	357	2	033169 mycobacteri
41	41	43.2	567	12	09DHS5 yaba-like d
42	41	43.2	597	2	09PR58 ureaplasma
43	41	43.2	658	2	09CP74 pasteurella
44	41	43.2	738	4	09P2P8 homo sapien
45	41	43.2	1146	2	051778 borrelia bu

## ALIGNMENTS

RESULT 1	
ID 069385	PRELIMINARY; PRT; 1390 AA.
AC 069385;	
DT 01-AUG-1998 (TREMBLrel. 07, Created)	
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)	
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE GLUCOSYLTRANSFERASE-SI.	
GN GTFc.	
OS Streptococcus mutans.	
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	
OC Streptococcus.	
OX NCBI_TaxID=1309;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN-M74245;	
RA MEDLINE=98231643; PubMed=9570124;	
RX Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.,	
RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans."	
RL EMBL: D88633; FAA26106.1; "	
DR EMBL: D88633; FAA26106.1; "	
DR InterPro: IPR002479; Glyco_hydro_70.	
DR InterPro: IPR003318; Glyco_hydro_70.	
DR Pfam: PF01473; CW_binding_1; 7.	
DR Pfam: PF02324; Glyco_hydro_70; 1.	
KW Transferase.	
SQ SEQUENCE 1390 AA; 155375 MW; 8847E4956FE05E9F CRC64;	
Query Match	95.8%; Score 91; DB 2; Length 1390;
Best Local Similarity	100.0%; Pred. No. 8.3e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY 1 VPSYFIRAHSEVODLI 18	
DB 578 VPSYFIRAHSEVODLI 595	
RESULT 2	
ID 069382	

ID 069382 PRELIMINARY; PRT; 1455 AA.  
AC 069382;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT8148;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88652; BAA26102.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162969 MW; 27D4D3A1EECA2939 CRC64;

Query Match 95.8%; Score 91; DB 2; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVODLI 18  
DB 578 VPSYSFIRAHSEVODLI 595

RESULT 3  
ID 069388 PRELIMINARY; PRT; 1455 AA.  
AC 069388;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4239;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88658; BAA26110.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 10.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match 95.8%; Score 91; DB 2; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVODLI 18  
DB 578 VPSYSFIRAHSEVODLI 595

RESULT 4  
ID 069391 PRELIMINARY; PRT; 1455 AA.  
AC 069391;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4251;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88661; BAA26114.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 95.8%; Score 91; DB 2; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVODLI 18  
DB 578 VPSYSFIRAHSEVODLI 595

RESULT 5  
ID 069397 PRELIMINARY; PRT; 1455 AA.  
AC 069397;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4467;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D89978; BAA26120.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.

SQ SEQUENCE 1455 AA; 162913 MW; A1263427BF24EBE1 CRC64;

Query Match 95.8%; Score 91; DB 2; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVODLI 18  
|||||

DB 578 VPSYSFIRAHDSVODLI 595

RESULT 6

O5263 PRELIMINARY; PRT; 1590 AA.

AC O5263;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE GRP-I.

OS Streptococcus sobrius.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OX NCBI\_TaxID=1310;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 33478;

RA Sato S.;

RL Amn. Kagoshima Univ. Dental School 16:23-29(1996).

DR EMBL; D63570; BAA09792.1; "

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_1; 15.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Transferase.

SQ SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 88.4%; Score 84; DB 2; Length 1590;  
Best Local Similarity 88.9%; Pred. No. 1.6e-05;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVODLI 18  
|||||

DB 548 VPSYSFIRAHDSVODII 565

RESULT 7

O5983 PRELIMINARY; PRT; 1590 AA.

AC O5983;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).

OS Streptococcus sobrius.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OX NCBI\_TaxID=1310;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OM2176;

RA MEDLINE=94146405; PubMed=8312602;

RA Sato S., Inoue M., Handa N., Aizawa Y., Isebe Y., Katayama T.;

RT "DNA sequence of the glucosyltransferase gene of serotype d Streptococcus sobrius.";

RL DNA Seq. 4:19-27(1993).

CC -I- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) - D-FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).

DR EMBL; D13858; BAA02976.1; "

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_1; 16.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Signal; Transferase; Glycosyltransferase.

FT SIGNAL 1 38 POTENTIAL.

FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I.

SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match 88.4%; Score 84; DB 2; Length 1590;  
Best Local Similarity 88.9%; Pred. No. 1.6e-05;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVODLI 18  
|||||

DB 548 VPSYSFIRAHDSVODII 565

RESULT 8

O91466 PRELIMINARY; PRT; 1477 AA.

AC O91466;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE DEXTRANSUCRASE (EC 2.4.1.5).

OS Leuconostoc mesenteroides.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

OX NCBI\_TaxID=1245;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NRL B-1355;

RA Arguello-Morales M.A., Remaud-Simeon M., Pizut S., Sarcabal P., Willemot R.M., Monsan P.;

RT "Sequence analysis of the gene encoding alternansucrase, a sucrose glucosyltransferase from Leuconostoc mesenteroides NRL B-1355.";

RL Submitted (CCF-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL; AJ250172; CAB76565.1; "

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_1; 14.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Transferase; Glycosyltransferase.

SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFC8B31 CRC64;

Query Match 84.2%; Score 80; DB 2; Length 1477;  
Best Local Similarity 73.7%; Pred. No. 7.1e-05;  
Matches 14; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVODLIA 19  
:|||||

DB 603 IPNYSFVRHDSVQTVIA 621

RESULT 9

O52224 PRELIMINARY; PRT; 1508 AA.

AC O52224;

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).

OS Leuconostoc mesenteroides.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

OX NCBI\_TaxID=1245;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-NRRL B-1299;  
 RA Monchois V., Remaud-Simeon M., Monsan P., Willemot R.M.,  
 RL FEKS Microbiol. Lett. 0:0-0(1998).  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
 FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
 DR EMBL: AF030129; AAB95453.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Transferrase; Glycosyltransferase.  
 KW SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 84.2%; Score 80; DB 2; Length 1508;  
 Best Local Similarity 73.7%; Pred. No. 7.2e-05;  
 Matches 14; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

1 VPSYFIRAHSEVODLIA 19  
 :|:||||:|||||||:|  
 DB 634 IPNYSFVRAHSEVQTVIA 652

RESULT 10  
 O9EZHS PRELIMINARY; PRT: 1508 AA.  
 AC O9EZHS:  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE DSRB742.  
 GN DSRB742.  
 OS Leuconostoc mesenteroides.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B-742CB;  
 RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
 RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene,"  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF294469; AAG38021.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 84.2%; Score 80; DB 2; Length 1508;  
 Best Local Similarity 73.7%; Pred. No. 7.2e-05;  
 Matches 14; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYFIRAHSEVODLIA 19  
 :|:||||:|||||||:|  
 DB 634 IPNYSFVRAHSEVQTVIA 652

RESULT 11  
 O9ZAR4 PRELIMINARY; PRT: 1527 AA.  
 AC O9ZAR4:  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 GN DEX.  
 OS Leuconostoc mesenteroides.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN (1)

RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B-512-F;  
 RA Bhattacharjee R., Singh D.K.S.;  
 RT "Cloning and Molecular Characterization of Dextranucrase Gene from  
 Leuconostoc mesenteroides NRRL B-512F.";  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U81374; AAD10952.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 16.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 84.2%; Score 80; DB 2; Length 1527;  
 Best Local Similarity 73.7%; Pred. No. 7.3e-05;  
 Matches 14; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYFIRAHSEVODLIA 19  
 :|:||||:|||||||:|  
 DB 652 IPNYSFVRAHSEVQTVIA 670

RESULT 12  
 O48756 PRELIMINARY; PRT: 1290 AA.  
 AC O48756:  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 OS Leuconostoc mesenteroides.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B1299;  
 RX MEDLINE-97136686; PubMed-8982063;  
 RA Monchois V., Willemot R.M., Remaud-Simeon M., Croux C., Monsan P.;  
 RT "Cloning and sequencing of a gene coding for a novel dextranucrase  
 from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-  
 6) and alpha (1-3) linkages.";  
 RL Gene 182:23-32(1996).  
 DR EMBL: U38181; AAB40875.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 11.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SEQUENCE 1290 AA; 145590 MW; 3555C2E96B749FAA CRC64;

Query Match 83.2%; Score 79; DB 2; Length 1290;  
 Best Local Similarity 83.3%; Pred. No. 9e-05;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 PYSYFIRAHSEVODLIA 19  
 :|:||||:|||||||:|  
 DB 388 IPNYSFVRAHSEVQTVIA 405

RESULT 13  
 O9LCJ7 PRELIMINARY; PRT: 1016 AA.  
 AC O9LCJ7:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 GN DSRT.  
 OS Leuconostoc mesenteroides.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.

OX NCBI\_TaxID=1245;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B-512F;  
 RA MEDLINE-20169623; PubMed-10705445;  
 RA Funane K., Mizuno K., Takahara H., Kobayashi M.;  
 RT "Gene encoding a dextranucrase-like protein in Leuconostoc  
 mesenteroides NRRL B-512F.";  
 RL Biosci. Biotechnol. Biochem. 64:29-38(2000).  
 DR EMBL; AB020020; BAA90527.1; -.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.1.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1016 AA; 110343 MW; 8896FEDE13CCB47 CRC64;

Query Match 81.1%; Score 77; DB 2; Length 1016;  
 Best Local Similarity 77.8%; Pred. No. 0.00015;  
 Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 PSYSFIRAHDSVODLIA 19  
 |||:|||||||:11  
 625 PNTYVFIRAHDSVOTVIA 642

RESULT 14  
 ID O9WXJ5 PRELIMINARY; PRT: 1512 AA.  
 AC O9WXJ5;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE GTF-S.  
 GN GTF.  
 OS Streptococcus criceti.  
 OC Plasmid pAM1.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1333;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HS-6;  
 RA Inoue M., Fukui K., Miyagi A.;  
 RT "S. cricetus glucosyltransferase(gtfs and gtfI) genes.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB026123; BAA77237.1; -.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 14.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Plasmid.  
 SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9C6C601FC14 CRC64;

Query Match 77.9%; Score 74; DB 2; Length 1512;  
 Best Local Similarity 77.8%; Pred. No. 0.00078;  
 Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 PSYSFIRAHDSVODLIA 19  
 |||:|||||||:11  
 560 PSYVFVIRAHDSVOTVIA 577

RESULT 15  
 ID O9WXJ4 PRELIMINARY; PRT: 1338 AA.  
 AC O9WXJ4;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE GTF-S.  
 GN GTF.  
 OS Streptococcus criceti.  
 OC Plasmid pAM1.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1333;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HS-6;  
 RA Inoue M., Fukui K., Miyagi A.;  
 RT "S. cricetus glucosyltransferase(gtfs and gtfI) genes.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB026123; BAA77236.1; -.  
 DR HSSP; P06278; IVUS.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 10.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Plasmid.  
 SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 76.8%; Score 73; DB 2; Length 1338;  
 Best Local Similarity 78.9%; Pred. No. 0.001;  
 Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVODLIA 19  
 |||:|||||||:11  
 509 VPNTYVFIRAHDSVOTRIA 527

Search completed: March 27, 2002, 14:26:02  
 Job time: 1674 sec









	Query Match	77.7%	Score 87:	DB 20:	Length 22:	
	Best Local Similarity	77.3%	Pred. No.	2.2e-08;		
	Matches 17;	Conservative	2;	Mismatches	3;	Indels 0; Gaps 0
Oy	1 TGARTINGQLTYFRANGVQVK 22					
Db	1 tgaqtikgqklyfkangqqvk 22					
RESULT 4						
ID AAW34164	AAW34164 standard; peptide; 24 AA.					
XX	AAW34164;					
AC	18-FEB-1998 (first entry)					
DT	GTF antigenic peptide #4 linked to polylysine core.					
DE						
XX						
KY	Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;					
●	surface domain; glucan-binding domain; mutans streptococcal strain;					
XX	immune response; glucan-binding activity; dental caries prevention.					
OS	Synthetic.					
OS	Streptococcus mutans.					
FH	Key	Location/Qualifiers				
FT	Peptide	1..22				
FT	/label=	GTF antigenic peptide #4 (see AAW34159)				
FT	/note=	"attached to the dendritic polylysine core via the alpha-amino group of Lys(23); a second copy of the antigenic 22-mer is linked to Lys(23) via the omega amino group"				
FT	Modified-site	23				
FT	/note=	"Lys(23) is linked to one copy of the antigenic peptide through the alpha-amino group, and to a second copy of the peptide (not shown) via the omega amino group"				
FT	Modified-site	24				
FT	/note=	"the alpha amino acid group of Lys(24) forms a peptide linkage with the carboxyl amino group of Lys(23); the omega amino group of Lys(24) forms a peptide bond with a second Lys residue analogous to Lys(23)"				
XX						
FN	US5686075-A.					
PD	11-NOV-1997.					
PD	01-MAY-1992;	92US-0877295.				
PR	30-APR-1993;	93US-0057162.				
PR	01-MAY-1992;	92US-0877295.				
XX	(FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.					
PA						
PI	Smith DJ, Taubman MA;					
XX						
DR	WPI; 1997-558089/51.					
XX						
PT	Immunogenic compositions containing streptococcal					
PT	glucosyl:transferase peptide(s) - used for provoking immune response					
PT	to streptococcal glucosyl:transferase for preventing dental caries					
PS	Claim 12; Column -; 11pp; English.					
CC						
CC	AAW34161-W34165 represent the Streptomyces mutans glycosyltransferase					
CC	(GTF) enzyme immunogenic fragments shown in AAW34155-W34160 linked to a					
CC	polylysine core. AAW34157 and AAW34158 are from the catalytic domain of					
CC	GTF, while AAW34160 is from the GTF surface domain. AAW34156 and AAW34155					
CC	are from the glucan-binding domain of GTF. These sequences, and the					
CC	immunogenic fragments shown in AAW34156-W34160 can all be used in the					

CC	immunogenetic composition of the invention. A composition comprising one of
CC	these sequences can be administered to a mammal to raise an immune
CC	response, in a method for interfering with the enzymatic activity of
CC	streptococcal glucosyltransferase in a mammal. The immune response
CC	results in reduction of the colonisation or accumulation of mutants
CC	streptococcal strains in the mammal. Compositions containing AAM34156
CC	specifically interfere with the glucan-binding activity of the
CC	streptococcal glucosyltransferase. The peptides can also be used in
CC	vaccines for preventing dental caries in mammals.
XX	
SQ	Sequence 24 AA:
OY	Query Match 77.7%; Score 87; DB 18; Length 24;
	Best Local Similarity 77.3%; Pred. No. 2.5e-08;
	Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
DB	1 TGARTINGOLLYFRANGVOVKG 22                  1 tgaqlikgklyfkangqvkxg 22
RESULT	5
ID	AAR32925
XX	AAR32925 standard; Protein; 1592 AA.
AC	AAR32925;
XX	
DT	28-JUN-1993 (first entry)
XX	
DE	Glucosyltransferase I.
XX	
KW	GT-1; Streptococcus; dental; caries.
XX	
OS	Streptococcus sobrinus.
XX	
PN	JP05023188-A.
XX	
PD	02-FEB-1993.
XX	
PF	25-JUL-1991; 9LJP-0186592.
XX	
PR	25-JUL-1991; 9LJP-0186592.
XX	
PA	(FUKU/) FUKU I.
PA	(KATO/) KATO K.
XX	
DR	WP1: 1993-079449/10.
DR	N-PSDB; AAQ37760.
XX	
PT	DNA sequence glucosyltransferase-I - comprises Streptococcus
PT	sobrinus DNA sequence with at least one nucleotide added or
PT	deleted
XX	
PS	Claim 13; "acc 15; 29pp; Japanese.
XX	
CC	The DNA sequence from Streptococcus sobrinus strain 6715 encodes
CC	glucosyltransferase-I (and mutants). The DNA was obtd. by treating
CC	S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,
CC	partially digesting with Sau3AI and fractionating on agarose gel.
CC	The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109
CC	transformed with it. A GT-1 expressing clone was isolated and
CC	sequenced. The clone may be used in the development of a drug for
CC	dental caries.
XX	
SQ	Sequence 1592 AA;
OY	Query Match 77.7%; Score 87; DB 14; Length 1592;
	Best Local Similarity 77.3%; Pred. No. 4.5e-06;
	Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY	1 TGARTINGOLLYFRANGVOVKG 22



```

XX OS Clostridium difficile.
XX PN MO9859053-A1.
XX PD 30-DEC-1998.
XX PF 19-JUN-1998; 98WO-GB01805.
XX PR 07-JAN-1998; 98GB-0000321.
XX PR 20-JUN-1997; 97GB-0013146.
XX PA (UNIO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX PA (QUEE-) QUEEN MARY & WESTFIELD COLLEGE.
XX PI Douce G, Dougan G, Ward SJ, Wren BW;
XX WPI; 1999-081281/07.
XX PT New isolated Clostridium difficile proteins - comprise fragments of
XX toxin A, used to develop agents for prophylaxis, treatment or
XX diagnosis of C. difficile infections
XX PS Claim 1; Fig 6; 82pp; English.
XX CC This sequence represents an immunogenic fragment of Clostridium difficile
XX (CD) Toxin A, designated fragment p5/6. This sequence is encoded by
XX nucleotides 7153-8118 of the toxin A gene. The invention relates to
XX fragments of toxin A of CD which are non-toxic and immunogenic. They can
XX be used in the preparation of vaccines against CD infection. The amino
XX acid molecules and antibodies against them can be used in the preparation
XX of an agent for the prophylaxis or treatment of a CD infection. They can
XX also be used for detection and diagnosis of CD infection.
XX SQ Sequence 320 AA;

Query Match 44.2%; Score 49.5; DB 20; Length 320;
Best Local Similarity 52.2%; Pred. No. 2;
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

OY 1 TGARTINGQLLYFRAN-GVOYKG 22
   |||||: || | | | |
Db 74 tglrtldgkkyfntntavtvg 96

RESULT 9
AAM87970
AAM87970 standard; peptide: 457 AA.

AAM87970:
XX 29-APR-1999 (first entry)
XX DT
XX DE Toxin A immunogenic fragment p5/7.
XX XX
XX KM Toxin A; immunogenic fragment; vaccine; diagnosis; infection.
XX OS Clostridium difficile.
XX PN MO9859053-A1.
XX PD 30-DEC-1998.
XX PF 19-JUN-1998; 98WO-GB01805.
XX PR 07-JAN-1998; 98GB-0000321.
XX PR 20-JUN-1997; 97GB-0013146.
XX PA (UNIO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX PA (QUEE-) QUEEN MARY & WESTFIELD COLLEGE.
XX PI Douce G, Dougan G, Ward SJ, Wren BW;

```

```

XX DR WPI; 1999-081281/07.
XX XX
XX PT New isolated Clostridium difficile proteins - comprise fragments of
XX toxin A, used to develop agents for prophylaxis, treatment or
XX diagnosis of C. difficile infections
XX PS Example 1; Fig 7; 82pp; English.
XX CC This sequence represents an immunogenic fragment of Clostridium difficile
XX (CD) Toxin A, designated fragment p5/7. This sequence is encoded by
XX nucleotides 6748-8118 of the Toxin A gene. The invention relates to
XX fragments of toxin A of CD which are non-toxic and immunogenic. They can
XX be used in the preparation of vaccines against CD infection. The amino
XX acid molecules and antibodies against them can be used in the preparation
XX of an agent for the prophylaxis or treatment of a CD infection. They can
XX also be used for detection and diagnosis of CD infection.
XX SQ Sequence 457 AA;

Query Match 44.2%; Score 49.5; DB 20; Length 457;
Best Local Similarity 52.2%; Pred. No. 3.1;
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

OY 1 TGARTINGQLLYFRAN-GVOYKG 22
   |||||: || | | | |
Db 211 tglrtldgkkyfntntavtvg 233

RESULT 10
AAR95014
AAR95014 standard; Protein: 811 AA.

AAR95014:
XX 08-JUL-1996 (first entry)
XX DT
XX DE C. difficile toxin A (aa1870-2680).
XX XX
XX KM Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;
XX diarrhoea; diagnosis; therapy.
XX OS Clostridium difficile VPI strain 10463 (ATCC 10463).
XX XX
XX W09612802-A1.
XX PD 02-MAY-1996.
XX PF 23-OCT-1995; 95WO-US13737.
XX PR 07-JUN-1995; 95US-0480604.
XX PR 24-OCT-1994; 94US-0329154.
XX PR 16-MAR-1995; 95US-0405496.
XX PR 14-APR-1995; 95US-0422711.
XX PA (OPHI-) OPHIDIAN PHARM INC.
XX PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
XX Williams JA;
XX WPI; 1996-230603/23.
XX PT Fusion proteins comprising non-toxin protein and part of toxin
XX useful to form anti-toxins against Clostridium botulinum type A, and
XX C. difficile type toxins, and to treat C. difficile intoxication,
XX partic. diarrhoea
XX PS Claim 53; Page 310-312; 434pp; English.
XX CC A recombinant protein (AAR95014) comprises amino acids 1870-2680 of
XX Clostridium difficile VPI strain 10463 toxin A (see also AAR95016),
XX a cytotoxin associated with diarrhoeic disease. It was obt'd. by

```



```

KW Toxin A repeating unit; rARU; toxin B repeating unit; rBRU; vaccine;
KM passive immune therapy; diarrhoea; colitis.
XX Clostridium difficile.
OS
FH Key Location/Qualifiers
FT Misc-difference 794 /note= "Gln encoded by GGTCAA"
FT Misc-difference 821..822 /note= "these residues encoded by GGA"
FT Misc-difference 836..866 /note= "the nucleotides encoding these residues are
FT not given"
XX
XX MOZ00061762-A1.
XX
XX 19-OCT-2000.
PD
XX
XX PF 10-APR-2000; 2000MO-USO9525.
XX
XX PF 09-APR-1999; 99US-0128686.
XX 01-MAR-2000; 2000US-0186201.
XX 20-MAR-2000; 2000US-0190111.
XX
XX (TECH-) TECHLABS INC.
XX
XX Wilkins TD, Lylerly DM, Moncrief JS, Zheng L, Phelps C;
XX WPI: 2000-679491/66.
XX DR N-PSDB: AAC62061.
XX
XX Immunogenic composition for use as a vaccine against Clostridium
PT difficile, comprises recombinant protein component consisting of at
PT least one repeating unit of toxin A or toxin B or both, of Clostridium
PT difficile
XX -
XX
XX Example 1; Fig 3; 39pp; English.
PS
CC The present sequence represents a Clostridium difficile recombinant
CC protein comprising toxin A repeating units (rARU). The protein is used
CC to prepare recombinant protein components. These recombinant protein
CC components comprise at least one repeating unit of Clostridium difficile
CC toxin A (rARU), one repeating unit of Clostridium difficile toxin B
CC (rBRU), or both. The recombinant protein is used as a vaccines, and is
CC useful for conferring a protective response in a mammalian host. It is
CC also useful for producing antibodies for passive immune therapy against
CC a strain of Clostridium difficile. The vaccines, formulated for use in
CC humans and animals, is useful for treating Clostridium difficile
CC mediated diarrhoea and colitis.
CC
Sequence 866 AA:

Query Match 44.2%; Score 49.5; DB 21; Length 866;
Best Local Similarity 52.2%; Pred. No. 6.9;
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 1 TGARTINGOLLYFRAN-GVQVKG 22
   |||||: | | | | |
Db 617 tglrtidgkkyfntntavvtg 639

RESULT 14
ID AAB30546
XX AAB30546 standard; Protein; 866 AA.
XX
XX AAB30546;
XX
XX 06-MAR-2001 (first entry)
XX
XX A recombinant protein of toxin A repeating units (rARU).
DE
XX
XX Toxin A; repeating unit; rARU; immunogenic composition; polysaccharide;

```

KW		pathogenic microorganism; nosocomial pathogenic microorganism; vaccine;
KM		Immune response.
XX		
OS	Clostridium difficile.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 794	/note= "Gln encoded by GGTCAA"
FT	Misc-difference 821..822	/note= "these residues encoded by GGA"
FT	Misc-difference 836..866	/note= "the nucleotides encoding these residues are not given"
FT		
XX		
PN	WO200061761-A2.	
PD	19-OCT-2000.	
XX		
PF	10-APR-2000; 2000WO-USO9523.	
XX		
PR	09-APR-1999; 99US-0128686.	
PR	01-MAR-2000; 2000US-0186201.	
XX		
PA	(TECH-) TECHLAB INC.	
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	
XX		
PI	Wilkins TD, Lylerly DM, Moncrief JS, Pavliakova D, Scheerson R;	
PI	Robbins JB;	
DR	WPI: 2000-679490/66.	
XX		
PT	Immunogenic compositions useful as vaccines comprise a recombinant protein of toxin A or B of Clostridium difficile conjugated to a polysaccharide of a microbial pathogen	-
PS	Example 1: Fig 3; 45pp; English.	
XX		
CC	The present sequence represents a Clostridium difficile recombinant protein comprising toxin A repeating units (rARU). The protein is used to prepare immunogenic compositions of the invention. The immunogenic compositions comprise a recombinant protein and a polysaccharide component, in which the recombinant protein is encoded by a gene from a strain of Clostridium difficile and the polysaccharide component is isolated from a strain of pathogenic microorganism or is chemically synthesized. The immunogenic compositions are useful for eliciting a protective immune response (T-cell dependent or T-cell independent, a cellular or humoral immune response) to a strain of pathogenic microorganism. The immunogenic composition also elicits a protective immune response against the polysaccharide produced by a strain of a nosocomial pathogenic microorganism. The immunogenic compositions are useful as vaccines for humans, particularly children and animals in affording protection against one or more microbial pathogens.	
SO	Sequence 866 AA:	
OY	Query Match 44.2%; Score 49.5; DB 21; Length 866; Best Local Similarity 52.2%; Pred. No. 6.9; Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;	
DB	1 TGARTINGOLLYFRAN-GVQVKG 22    ::   ::    617 tglrtldgkkyfntnfavtvg 639	
RESULT 15		
ID	AAR95016	
XX	AAR95016 standard; Protein: 2710 AA.	
XX	AAR95016;	
DT	08-JUL-1996 (first entry)	
XX		

DE C. difficile toxin A.

XX  
KW Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;  
KM vaccine; diarrhoea; diagnosis; therapy.

XX  
OS Clostridium difficile VPI strain 10463 (ATCC 10463).

XX  
PN W09612802-A1.

XX  
PD 02-MAY-1996.

XX  
PF 23-OCT-1995; 95WO-US13737.

XX  
PR 07-JUN-1995; 95US-0480604.

PR 24-OCT-1994; 94US-0329154.

PR 16-MAR-1995; 95US-0405496.

PR 14-APR-1995; 95US-0422711.

XX  
PI (OPHI-) OPHIDIAN PHARM INC.

PI Firca JR, Klink JA, Padhye NV, Stafford DC, Thalley BS.

XX  
PI Williams JA;

XX  
DR WPI; 1996-230603/23.

XX  
N-PSDB; AAT29248.

XX  
PT Fusion proteins comprising non-toxin protein and part of toxin

PT useful to form anti-toxins against Clostridium botulinum type A, and

PT C. difficile type toxins, and to treat C. difficile intoxication,

PT partic. diarrhoea

PS Claim 63; Page 290-302; 434pp; English.

XX  
CC Clostridium difficile VPI strain 10463 toxin A (AAR95016), the

CC product of the toxin A gene (AAT29248), is a potent cytotoxin that

CC plays a direct role in damaging gastrointestinal tissues and is

CC associated with diarrhoeic disease. It can be obtd. by expression in

CC transformed E. coli hosts of portions of DNA that together cover the

CC entire toxin A gene. Toxin A, and portions of it (see also

CC AAR95014-15 and AAR95017), pref. expressed as fusions to polystyridine

CC affinity tags or maltose binding protein, are used to raise avian

CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.

XX  
SQ Sequence 2710 AA;

XX  
Query Match 44.2%; Score 49.5; DB 17; Length 2710;

XX  
St Local Similarity 52.2%; Pred. No. 29;

XX  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 1 TGARTINGOLLYFRAN-GYQVKG 22

DB 2460 tglrtldgkkyfntnctavavcg 2482

Search completed: March 27, 2002, 13:57:54  
Job time: 522 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:01 ; Search time 53.4 Seconds  
(without alignments)  
15.105 Million cell updates/sec

```
Title: US-09-290-049A-3
Perfect score: 112
Sequence: 1 TGARTINGQLLYFRANGVQVKG 22
```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

```

Maximum DB seq length: 0
Maximum DB seq length: 20000000000

```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

Database :        SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	DB	ID	Description
1	105	93.8	1476	1	GTFB_STPMU	P06887 streptococc
2	87	77.7	1592	1	GTFB_STPMU	P27470 streptococc
3	87	77.7	1597	1	GTFB_STPMU	P11001 streptococc
4	78	69.6	1375	1	GTFB_STPMU	P13470 streptococc
5	75	67.0	1365	1	GTFB_STPMU	P29336 streptococc
6	67	59.8	1462	1	GTFB_STPMU	P49331 streptococc
7	49.5	44.2	2710	1	Y119_NPPOP	O10358 orygia pseu
8	49.5	44.2	2710	1	TOXA_CLODI	P16154 clostridium
9	47	42.0	785	1	DOC_BPPI	O06259 bacteriophu
10	47	42.0	788	1	SPL_HUMAN	P08047 bacteriophu
11	47	42.0	231	1	SPL_RAT	O01714 ratus norv
12	46	41.1	231	1	DLHM_AQUAE	O67802 aquifex aeo
13	46	41.1	245	1	USF_AQUAE	P46209 aquifex pyr
14	46	41.1	301	1	T2M4_MERJA	O58723 methanococc
15	46	41.1	1396	1	LECI_MOUSE	P24721 mus musculu
16	44	39.3	110	1	VLTE_BPPI	P13390 bacteriophu
17	43	38.4	299	1	YC27_ARCFU	O29041 archaeoglob
18	43	38.4	467	1	YC02_CARFL	O09529 caenorhabd
19	43	38.4	467	1	ISPE_SCHPO	P40903 schizosacch
20	41.5	37.1	401	1	ISPE_SCHPO	P40903 schizosacch
21	41.5	37.1	472	1	HAT2_YEAST	P393884 saccharomyc
22	41.5	37.1	472	1	CBIR_FELCA	O02777 felis silve
23	41.5	37.1	472	1	CBIR_HUMAN	P21554 homo sapien
24	41.5	37.1	473	1	CBIR_MOUSE	P47746 mus musculu
25	41.5	37.1	473	1	CBIR_RAT	P20772 ratus norv
26	41.5	37.1	567	1	DY13_CHIRE	P27766 chlamydomon
27	41	36.6	579	1	COE2_BRARE	O93375 brachydanio
28	41	36.6	336	1	VINT_BP186	O93375 brachydanio
29	41	36.6	412	1	YQW_BACSU	P06723 bacteriussu
30	41	36.6	433	1	MURC_PORGI	P54560 bacillus su
31	41	36.6	473	1	CBIR_TARGA	O51831 porphyromon
32	40.5	36.2	464	1	Y073_TREPA	O09pi17 taricha gra
33	40.5	36.2	524	1	VL2_HPV27	O83112 treponema p
					VL2_HPV2A	P36755 human papill
						P25487 human papill

	34	40	35.7	394	1	NDOC_SYNV3	P27724 synechocyst
	35	40	35.7	460	1	ALN_YEAST	P32375 saccharomyc
	36	40	35.7	467	1	REAL_CRIFA	Q33696 citrithidia f
	37	40	35.7	691	1	TOP1_BACSU	P39814 bacillus su
	38	40	35.7	1258	1	GLI2_HUMAN	P10070 homo sapien
	39	39.5	35.3	473	1	CBIR_PORCU	P56971 poephilla gu
	40	39	34.8	301	1	LEC1_RAT	P08290 rattus norv
	41	39	34.8	310	1	YES9_PYRHO	O59128 pyrococcus
	42	39	34.8	333	1	MAUG_MERFL	Q50426 methyllobaci
	43	39	34.8	382	1	YJ07_YEAST	P47000 saccharomyc
	44	39	34.8	402	1	YIN2_STRAM	P32426 streptomyc
	45	39	34.8	404	1	CAG5_CHICK	G92184 gallus gall

ALIGNMENTS

```

RESULT_1
GTFB_STRMU STANDARD; PRT; 1476 AA.
ID GTFB_STRMU STANDARD; PRT; 1476 AA.
AC P08987; 069381; 069384; 069387; 069390; 069396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)
GN GTFB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
ON NCBI_TaxId=1309;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfb gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC SPRAIN-MT4239, MT4245, MT4251, MT4467, AND MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans".
RL FEMS Microbiol. Lett. 161:331-336(1998).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBR'S.
CC -1- CATALYTIC ACTIVITY: SUCCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE** + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
CC CC
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE HTTP://WWW.ISB-CH/ANNOUNCE/
CC OR SEND AN EMAIL TO LICENSE@ISB-CH).
CC CC
CC EMBL, M17361, AAA86588.1, -
CC EMBL, D88651, BAA26101.1, -
CC EMBL, D88654, BAA26105.1, -

```

```

DR EMBL: D88657; BAA26109.1; -
DR EMBL: D88660; BAA26113.1; -
DR EMBL: D89977; BAA26119.1; -
DR PIR: B3135; B3135.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro.70.
DR Pfam: PF01473; CW_binding.1; 13.
DR Pfam: PF02324; Glyco_hydro.70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1
FT CHAIN 35 1476
FT DOMAIN 35 1051
FT DOMAIN 1097 1476
FT REPEAT 1097 1130
FT REPEAT 1161 1470
FT REPEAT 1161 1210
FT REPEAT 1225 1275
FT REPEAT 1290 1340
FT REPEAT 1355 1405
FT REPEAT 1420 1470
FT REPEAT 62 62
FT VARIANT 62 65
FT VARIANT 65 65
FT VARIANT 68 68
FT VARIANT 78 78
FT VARIANT 86 86
FT VARIANT 89 89
FT VARIANT 168 168
FT VARIANT 276 276
FT VARIANT 399 399
FT VARIANT 474 474
FT VARIANT 512 512
FT VARIANT 519 519
FT VARIANT 701 701
FT VARIANT 708 708
FT VARIANT 938 938
FT VARIANT 952 957
FT VARIANT 963 964
FT VARIANT 968 970
FT VARIANT 1086 1086
FT VARIANT 1158 1158
FT VARIANT 1163 1163
FT VARIANT 1168 1168
FT VARIANT 1182 1182
FT VARIANT 1234 1234
FT VARIANT 1263 1263
FT VARIANT 1263 1263
FT VARIANT 1264 1264
FT VARIANT 1272 1272
FT VARIANT 1329 1329
FT VARIANT 1394 1394
FT VARIANT 1402 1402
FT VARIANT 1459 1459
FT VARIANT 570 570
FT CONFLICT 800 817
FT CONFLICT 1310 1310
FT SEQUENCE 1476 AA; 165685 MW; 3479B62B07694D98 CRC64;

Query Match          93.8%; Score 105; DB 1; Length 1476;
Best Local Similarity 95.5%; Pred. No. 7.5e-09;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

RESULT 2
ID GTF2_STRDO STANDARD; PRT; 1592 AA.
AC GTF2_STRDO
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)
DE (SUCCROSE 6-GLUCOSYLTRANSFERASE).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=1317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6715;
RX MEDLINE=91123227; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL Bacteriol. 173:989-996(1991).
CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS. THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -I- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
CC -I- D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N)+1.
CC -I- SUBCELLULAR LOCATION: SECRETED.
CC -I- DISEASE: DENTAL CARIES.
CC -I- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as there is no content in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: D90213; BAA14241.1; -
DR PIR: A38175; A38175.
DR HSP: P00695; 2HRE.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro.70.
DR Pfam: PF01473; CW_binding.1; 16.
DR Pfam: PF02324; Glyco_hydro.70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1
FT CHAIN 39 1592
FT DOMAIN 39 1044
FT DOMAIN 1093 1592
FT DOMAIN 1093 1592
FT REPEAT 1093 1142
FT REPEAT 1158 1207
FT REPEAT 1222 1272
FT REPEAT 1287 1337
FT REPEAT 1402 1451
FT REPEAT 1514 1563
FT REPEAT 1577 1592
FT SEQUENCE 1592 AA; 176167 MW; BC0A66D079351ECF CRC64;

Query Match          77.7%; Score 87; DB 1; Length 1592;
Best Local Similarity 77.3%; Pred. No. 6.2e-06;

```



FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1126 1375 2.4 A, 1 C AND 1 AC REPEATS.  
 FT REPEAT 1126 1159 A REPEAT.  
 FT REPEAT 1169 1200 A REPEAT.  
 FT REPEAT 1227 1238 C REPEAT.  
 FT REPEAT 1253 1303 AC REPEAT.  
 FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).  
 FT REPEAT 1375 1330 A REPEAT (INCOMPLETE).  
 SQ SEQUENCE 1375 AA; 153022 MW; D4B80CBEE0A0ACE13 CRC64;

Query Match 69.6%; Score 78; DB 1; Length 1375;  
 Best Local Similarity 68.2%; Pred. No. 0.00015;  
 Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGARTINGQLYFRANGVOYKG 22  
 1264 TGIVTFNGRLYFKPNGVOYAKG 1285

RESULT 5  
 GTF5\_STRDO STANDARD; PRT: 1365 AA.  
 ID GTF5\_STRDO  
 AC P29336;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
 DE (SUROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTF5.  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ME28;  
 RA MEDLINE=90316665; PubMed=2142479;  
 RX Gilmore K.S., Russell R.R., Ferretti J.J.;  
 RT "Analysis of the Streptococcus downei gtfS gene, which specifies a  
 glucosyltransferase that synthesizes soluble glucans.";  
 RL Infect. Immun. 58:2452-2458(1990).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
 PRIMER GLUCAN UNLIKE GTF-I.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA  
 1,6-GLUCOSE).  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 CC -----  
 CC This SMSS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M30943; AAA26898.1; -  
 DR PIR: A41483; A41483.  
 DR InterPro: IPR002479; CM\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CM\_binding\_1; 10.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1  
 FT CHAIN 37 1365 OR 37 (POTENTIAL).  
 FT GLUCOSYLTRANSFERASE-S.

FT DOMAIN 37 1050 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1083 1365 4.5 X TANDEM REPEATS.  
 FT REPEAT 1083 1131 1.  
 FT REPEAT 1150 1199 2.  
 FT REPEAT 1225 1274 3.  
 FT REPEAT 1289 1339 4.  
 FT REPEAT 1353 1365 5 (INCOMPLETE).  
 SQ SEQUENCE 1365 AA; 151590 MW; 167296B5A2E8C476 CRC64;

Query Match 67.0%; Score 75; DB 1; Length 1365;  
 Best Local Similarity 63.6%; Pred. No. 0.00044;  
 Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGARTINGQLYFRANGVOYKG 22  
 DB 1235 TGEOTIDGCKVEFDNGVOYKG 1256

RESULT 6  
 GTFD\_STRMU STANDARD; PRT: 1462 AA.  
 ID GTFD\_STRMU  
 AC P49331; O69383; O69386; O69389; O69392; O69398;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
 DE (SUROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFD.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GS-5;  
 RA MEDLINE=9110958; PubMed=2148600;  
 RX Honda O., Kato C., Kuramitsu H.K.;  
 RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
 the glucosyltransferase-S enzyme.";  
 RL J. Gen. Microbiol. 136:2099-2105(1990).  
 RN [2].  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MT429; MT4245; MT4251; MT4467; AND MT8148;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
 RA Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 RL FEMS Microbiol. Lett. 161:331-336(1998).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 FORMS OF GLUCANS.  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 CC -----  
 CC This SMSS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC	EMBL: M29296; AAA26895.1; -	Score 67; DB 1; Length 1462;	59.8%;
DR	EMBL: D88653; BAA26103.1; - <td></td> <td></td>		
DR	EMBL: D88656; BAA26107.1; - <td></td> <td></td>		
DR	EMBL: D88659; BAA26111.1; - <td></td> <td></td>		
DR	EMBL: D88662; BAA26115.1; - <td></td> <td></td>		
DR	EMBL: D89979; BAA26121.1; - <td></td> <td></td>		
DR	InterPro: IPR002479; Cw_binding. <td></td> <td></td>		
DR	InterPro: IPR003318; Glyco_hydro.70. <td></td> <td></td>		
DR	Pfam: PF01473; Cw_binding_1; 11. <td></td> <td></td>		
DR	Pfam: PF02324; Glyco_hydro.70; 1. <td></td> <td></td>		
KW	Transferase; Glycosyltransferase; Signal; Repeat; Dental caries. <td></td> <td></td>		
FT	SIGNAL 1 ? <td></td> <td></td>		
FT	CHAIN ? 1462 <td></td> <td></td>		
FT	DOMAIN 1332 1423 <td></td> <td></td>		
FT	REPEAT 1232 1295 <td></td> <td></td>		
FT	REPEAT 1296 1359 <td></td> <td></td>		
FT	REPEAT 1360 1423 <td></td> <td></td>		
FT	VARIANT 58 58 <td></td> <td></td>		
FT	VARIANT 68 68 <td></td> <td></td>		
FT	VARIANT 81 81 <td></td> <td></td>		
FT	VARIANT 113 113 <td></td> <td></td>		
FT	VARIANT 122 122 <td></td> <td></td>		
FT	VARIANT 132 132 <td></td> <td></td>		
FT	VARIANT 135 135 <td></td> <td></td>		
FT	VARIANT 202 202 <td></td> <td></td>		
FT	VARIANT 255 255 <td></td> <td></td>		
FT	VARIANT 275 275 <td></td> <td></td>		
FT	VARIANT 288 288 <td></td> <td></td>		
FT	VARIANT 301 301 <td></td> <td></td>		
FT	VARIANT 313 313 <td></td> <td></td>		
FT	VARIANT 317 317 <td></td> <td></td>		
FT	VARIANT 328 328 <td></td> <td></td>		
FT	VARIANT 350 350 <td></td> <td></td>		
FT	VARIANT 628 633 <td></td> <td></td>		
FT	VARIANT 688 688 <td></td> <td></td>		
FT	VARIANT 726 732 <td></td> <td></td>		
FT	VARIANT 726 730 <td></td> <td></td>		
FT	VARIANT 964 964 <td></td> <td></td>		
FT	VARIANT 1019 1019 <td></td> <td></td>		
FT	VARIANT 1059 1060 <td></td> <td></td>		
FT	VARIANT 1060 1060 <td></td> <td></td>		
FT	VARIANT 1080 1080 <td></td> <td></td>		
FT	VARIANT 1142 1142 <td></td> <td></td>		
FT	VARIANT 1198 1198 <td></td> <td></td>		
FT	VARIANT 1220 1220 <td></td> <td></td>		
FT	VARIANT 1280 1280 <td></td> <td></td>		
FT	VARIANT 1282 1282 <td></td> <td></td>		
FT	VARIANT 1290 1290 <td></td> <td></td>		
FT	VARIANT 1311 1311 <td></td> <td></td>		
FT	VARIANT 1403 1403 <td></td> <td></td>		
FT	VARIANT 1425 1425 <td></td> <td></td>		
FT	VARIANT 1449 1449 <td></td> <td></td>		
FT	CONFLICT 1428 1462 <td></td> <td></td>		
FT	SEQUENCE 1462 AA; 163512 MW; 56541IPDCHBDP00 CRC64; <td></td> <td></td>		
Query Match	Best Local Similarity 59.18%;		
Matches 13; Conservative 5; Mismatches 4; Indels 0; Gaps 0;			
1 TGARTINGQLYFRANGVQYKG 22			
:     :     :			

```

Db      1332  TGSQTIACKTLYFASDGQVKG 1353

RESULT      7
Y119_NPVOF  STANDARD;          PRT;    529 AA.
AC      010358:
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      HYPOTHEETICAL 59.0 KDA PROTEIN PRECURSOR (OBF119).
OS      Orygia pseudotsugata multinnucleocapsid virus (OPMNPV).
OC      Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC      Nucleopolyhedrovirus.
OX      NCBI_TaxID=164623;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97271300; PubMed=9126251;
RA      Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA      Rohrmann G.F.;
RT      "The sequence of the Orygia pseudotsugata multinnucleocapsid nuclear
RT      polyhedrosis virus genome."
RL      Virology 229:381-399(1997).
CC      -I- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on ways
CC      CC use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U75930; AAC59118.1; -
KM      Hypothetical protein; Signal.
FT      SIGNAL 1 20 POTENTIAL.
FT      CHAIN 21 529 HYPOTHEETICAL PROTEIN ORF119.
SQ      SEQUENCE 529 AA; 38983 MW; E85F81DD9219BCB CRC64;

Query Match      44.2%; Score 49.5; DB 1; Length 529;
Best Local Similarity 50.0%; Pred. No. 2;
Matches 11; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

QY      1 TGARTINGQLLYFRANGQVKG 22
      :||:|:| | :||:| |
Db      260  SGRRT-SCRLTFYHAADGVEVSG 280

RESULT      8
TOXA_CLODI  STANDARD;          PRT;    2710 AA.
ID      TOXA_CLODI
AC      P16154:
DT      01-APR-1990 (Rel. 14, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      TOXIN A.
DE      TOXA OR TCDA.
GN      Clostridium difficile.
OS      Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC      Clostridium.
OC      Clostridium.
OX      NCBI_TaxID=1496;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=VPI 10463;
RX      MEDLINE=90221894; PubMed=2109310;
RA      Sauerborn M., von Eichel-Streiber C.;
RT      "Nucleotide sequence of Clostridium difficile toxin A."
RL      Nucleic Acids Res. 18:1629-1630(1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=VPI 10463;
RC      MEDLINE=90129305; PubMed=2105276;
RX      MEDLINE=90129305; PubMed=2105276;
```

RA Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Lively D.M.,  
 RA Wilkins T.W., Johnson J.L.:  
 RT "Molecular characterization of the Clostridium difficile toxin A  
 RL gene." Infect. Immun. 58:480-488(1990).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-VPI 10463:  
 RA von Eichel-Streiber C.:  
 RL Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA  
 CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE  
 CC DIFFERENT OLIGOPEPTIDES.  
 CC -1- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN  
 CC ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL  
 CC DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE  
 CC CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X51797; CAA36094.1; -  
 DR EMBL: M30307; AAA23283.1; -  
 DR EMBL: X92982; CAA63564.1; -  
 DR PIR: S08638; S08638.  
 DR InterPro: IPR002479; CW\_binding.  
 DR Pfam: PF01473; CW\_binding\_1; 31.  
 DR Toxin: Enterotoxin.  
 KM SEQUENCE 2710 AA; 308052 MW; 0A6E52CE84C14421 CRC64;  
 SQ

Query Match 44.2%; Score 49.5; DB 1; Length 2710;  
 Best Local Similarity 52.2%; Pred. No. 11;  
 Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 1 TGARTINGQLYFRAN-GVQVKG 22

DB 2460 TGLRTIDGRKRYFNTNTAVAVTG 2482

RESULT 9  
 BPPI DOC\_BPPI STANDARD: PRT: 126 AA.  
 AC 006259;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE DEATH ON CURING PROTEIN.  
 GN DOC.  
 OS Bacteriophage P1.  
 OC VIRUSES; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.  
 OX NCBI\_TaxID=10678;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA MEDLINE=94016561; PubMed=8411153;  
 RA Lehnher H., Maguin E., Jafri S., Yarmolinsky M.B.:  
 RT "Plasmid addition genes of bacteriophage P1: doc, which causes cell  
 RT death on curing of prophage, and phd, which prevents host death when  
 RT prophage is retained." J. Mol. Biol. 233:414-428(1993).  
 RL J. Mol. Biol. 233:414-428(1993).  
 CC -1- FUNCTION: KILLS CELLS. DOC AND PHD PROTEIN FUNCTION IN UNISSON TO  
 CC STABILIZE PLASMID NUMBER BY INDUCING A LETHAL RESPONSE TO PLASMID  
 CC LOSS.  
 CC -1- MISCELLANEOUS: THE CONCENTRATION OF PHD IN P1 LYSOGENS IS FAR  
 CC GREATER THAN THAT OF THE POISON IT ANTAGONIZES, SUCH AN EXCESS MAY  
 CC ASSUME THE WELL-BEING OF CARRIERS OF THE ADDICTING PLASMID.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M95666; AAA16931.1; -  
 DR PIR: S38555; S38555.  
 DR SEQUENCE 126 AA; 13588 MW; E04A397538A595CD CRC64;  
 SQ

Query Match 42.0%; Score 47; DB 1; Length 126;  
 Best Local Similarity 60.0%; Pred. No. 1.1;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 6 INGQLLYFRANGVQV 20

DB 77 LNSALLFLRNGVQV 91

RESULT 10  
 ID SPl\_HUMAN  
 AC P08047; Q9NYE7; Q9H305;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE TRANSCRIPTION FACTOR SPl.  
 GN SPl OR TSEPl.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE OF 4-785 FROM N.A.  
 RC TISSUE=Cervical carcinoma;  
 RA Haggart M.H., Ladurner A.G.;  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE OF 1-558 FROM N.A.  
 RP MEDLINE=20545561; PubMed=10973950;  
 RA Takahara T., Kanazu S., Yanagisawa S., Akanuma H.;  
 RT "Heterogeneous SPl mRNAs in human HepG2 cells include a product of  
 RT homotypic trans-splicing." J. Biol. Chem. 275:38067-38072(2000).  
 RN [3]  
 RN SEQUENCE OF 90-785 FROM N.A., AND SEQUENCE OF 359-375 AND 670-675.  
 RP MEDLINE=88080466; PubMed=3319186;  
 RA Kadohaga J.T., Carner K.R., Maslarsz F.R., Tjian R.;  
 RT "Isolation of cDNA encoding transcription factor SPl and functional  
 RT analysis of the DNA binding domain." Cell 51:1079-1090(1987).  
 RL Cell 51:1079-1090(1987).  
 RN [4]  
 RN O-GLYCOSYLATION.  
 RP MEDLINE=89003041; PubMed=3139301;  
 RA Jackson S.P., Tjian R.;  
 RT "O-glycosylation of eukaryotic transcription factors: implications  
 RT for mechanisms of transcriptional regulation." Cell 55:125-133(1988).  
 RN [5]  
 RN STRUCTURE BY NMR OF 654-684 AND 684-712.  
 RP MEDLINE=97218212; PubMed=9065444;  
 RA Narayan V.A., Kiriacki R.W., Cardona J.P.;  
 RT "Structures of zinc finger domains from transcription factor SPl.  
 RT Insights into sequence-specific protein-DNA recognition." J.  
 RL J. Biol. Chem. 272:7801-7809(1997).  
 RN [6]  
 RN IDENTIFICATION OF SEROTONIN 1A RECEPTOR PROMOTER BINDING SITES.  
 RP MEDLINE=96224025; PubMed=8626793;  
 RA Parks C.L., Shenk T.;  
 RT "The serotonin 1a receptor gene contains a TATA-less promoter that  
 RT responds to MAZ and SPl.";

```
RL J. Biol. Chem. 271:4417-4430(1996).
CC -1- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY
CC ACTIVATES MRNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL
CC RECOGNITION SITES. CAN INTERACT WITH G/C-RICH MOTIFS FROM
CC SEROTONIN RECEPTOR PROMOTER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PPM: O-GLYCOSYLATED; CONTAINS N-ACETYLGLUCOSAMINE SIDE CHAINS.
CC -1- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF252284; AAF67726.1; -.
CC EMBL: AB039286; BAB13476.1; -.
CC PIR: A29635; A29635.
CC PDB: 1SP1; 2I-APR-97.
CC TRANSFAC: T00759; -.
CC GLYCOSULEDB: P08047; -.
CC MIM: 189906; -.
CC InterPro: IPR000822; Znf-C2H2.
CC Pfam: PF00096; Zf-C2H2; 3.
CC SMART: SM00355; Znf-C2H2; 3.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
CC KW DNA-binding; Nuclear protein; Repeats; Glycoprotein; 3D-structure.
CC KM DNA-binding; Nuclear protein; Repeat; Glycoprotein; 3D-structure.
CC FT ZN_FING 626 708 ZINC FINGERS.
CC FT ZN_FING 626 650 C2H2-TYPE.
CC FT ZN_FING 656 680 C2H2-TYPE.
CC FT ZN_FING 686 708 C2H2-TYPE.
CC FT CONFLICT 366 366 D -> G (IN REF. 3; AA SEQUENCE).
CC FT CONFLICT 670 670 S -> P (IN REF. 3; AA SEQUENCE).
CC FT SEQUENCE 785 AA; 80693 MW; 43893DBF6518B9EA CRC64;
SQ
Query Match 42.0%; Score 47; DB 1; Length 785;
Best Local Similarity 50.0%; Pred. No. 7.6;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
OY 4 RTINGQLLYFRANGVQVK 21
:|:|:| | | | | | | | | | |
180 QTVDSGQLQFATGAGVQ 197
RESULT 11
SPL_RAT ID SPL_RAT STANDARD; PRT; 788 AA.
AC 001714;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSCRIPTION FACTOR SPL.
GN SPL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93010958; PubMed=1356762;
RA Inatake H., Sogawa K., Yasumoto K., Kikuchi Y., Sasano K.,
RA Kobayashi A., Hayami M., Fujii-Kuriyama Y.;
RT "Two regulatory proteins that bind to the basic transcription element
(BTE), a GC box sequence in the promoter region of the rat p-450IA1
```

```
RT gene.";
RL EMBL J. 11:3663-3671(1992).
CC -1- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY
CC ACTIVATES MRNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL
CC RECOGNITION SITES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PPM: O-GLYCOSYLATED; CONTAINS N-ACETYLGLUCOSAMINE SIDE CHAINS
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: D12768; BAA02235.1; -.
CC PIR: JS0747; JS0747.
CC HSSP: P08047; 1SP1.
CC TRANSFAC: T00754; -.
CC InterPro: IPR000822; Znf-C2H2.
CC Pfam: PF00096; Zf-C2H2; 3.
CC PRINTS: PR00048; ZINC_FINGER.
CC SMART: SM00355; Znf-C2H2; 3.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
CC KW DNA-binding; Nuclear protein; Repeats; Glycoprotein.
CC KM DNA-binding; Nuclear protein; Repeat; Glycoprotein.
CC FT ZN_FING 629 711 ZINC FINGERS.
CC FT ZN_FING 629 653 C2H2-TYPE.
CC FT ZN_FING 659 683 C2H2-TYPE.
CC FT ZN_FING 689 711 C2H2-TYPE.
CC FT SEQUENCE 788 AA; 81015 MW; AA2B0CAB81AAB80C CRC64;
SQ
Query Match 42.0%; Score 47; DB 1; Length 788;
Best Local Similarity 50.0%; Pred. No. 7.7;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
OY 4 RTINGQLLYFRANGVQVK 21
:|:|:| | | | | | | | | | |
Db 183 QTVDSGQLQFATGAGVQ 200
RESULT 12
DLHH_AQUAE ID DLHH_AQUAE STANDARD; PRT; 231 AA.
AC 067802;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE CARBOXYMETHYLENEBUTENOLIDASE (EC 3.1.1.45) (DIENELACTONE
DE HYDROLASE) (DLH).
GN AQ_1997.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjey M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus."
RL Nature 392:353-358(1998).
CC -1- CATALYTIC ACTIVITY: 4-CARBOXYMETHYLENEBUT-2-EN-4-OLIDE + H(2)O =
CC 4 OXOHED-2-EMEDIOATE.
CC -1- SIMILARITY: BELONGS TO THE DIENELACTONE HYDROLASE FAMILY.
```

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; AE000767; AAC07773.1; -.  
 DR InterPro: IPR002925; DLH.  
 DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 DR Pfam; PF01738; DLH; 1.  
 KM Hypothetical protein: Hydrolase; Complete proteome.  
 FT ACT\_SITE 118 118 BY SIMILARITY.  
 FT ACT\_SITE 167 167 BY SIMILARITY.  
 FT ACT\_SITE 199 199 BY SIMILARITY.  
 FT SEQUENCE 231 AA; 26357 MW; 8268DA32A6980751 CRC64;

Query Match 41.1%; Score 46; DB 1; Length 231;  
 Best Local Similarity 53.3%; Pred. No. 3.1;  
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 8 GOLLYFRANGVQVKG 22  
 ||:| ||:|:|  
 Db 2 GQWVEFKNGKVGKRG 16

RESULT 13  
 ID USF\_AOUPY STANDARD: PRT; 231 AA.  
 AC P46209;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE USF PROTEIN.  
 GN USF.  
 OS Aquifex pyrophilus.  
 CC Bacteria; Aquificales; Aquificaceae; Aquifex.  
 CX NCBI\_TaxID-2714;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-KO15A;  
 RX MEDLINE-96062250; PubMed-7592443;  
 RA Benamer W., Shao Z., Mages W., Rachel R., Stetter K.O.,  
 RA Schmitt R.;  
 RI "Flagellar structure and hyperthermophily: analysis of a single  
 RI flagellin gene and its product in Aquifex pyrophilus.";  
 RL J. Bacteriol. 177:6630-6637(1995).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; U17575; AAA8922.1; -.  
 DR InterPro: IPR002925; DLH.  
 DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 DR Pfam; PF01738; DLH; 1.  
 FT SEQUENCE 231 AA; 26140 MW; C55E40839920195 CRC64;

Query Match 41.1%; Score 46; DB 1; Length 231;  
 Best Local Similarity 53.3%; Pred. No. 3.1;  
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 8 GOLLYFRANGVQVKG 22  
 ||:| ||:|:|  
 Db 2 GQWVEFKNGKVGKRG 16

RESULT 14  
 ID T2M4\_METJA STANDARD: PRT; 245 AA.  
 AC O58723;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE TYPE II RESTRICTION ENZYME MJAIIV (EC 3.1.21.4) (ENDONUCLEASE MJAIIV)  
 DE (R. MJAIIV).  
 GN MJAIIV OR MJ1327.  
 OS Methanococcus jannaschii.  
 CC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
 CC Methanococcus  
 CC NCBI\_TaxID-2190;  
 CX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE-96337999; PubMed-8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Feldman J.F., Fuhmann J.L., Nguyen D.,  
 RA Ullrichback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RA jannaschii.";  
 RT Science 273:1058-1073(1996).  
 RN [2]  
 RP CHARACTERIZATION.  
 RA Noren C.J., Roberts R.J., Patel J., Byrd D.R., Morgan R.D.;  
 RT "Method for screening restriction endonucleases.";  
 RL Patent number WO9911821, 11-MAR-1999.  
 CC -1- CATALYTIC ACTIVITY: RECOGNIZES THE DOUBLE-STRANDED UNMETHYLATED  
 CC SEQUENCE GTNNAC.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; U67573; AAB99343.1; -.  
 DR REBASE; 3894; MJaiV.  
 DR TIGR; MJ1327; -.  
 KM Hydrolyase; Endonuclease; Nuclease; Restriction system;  
 KW Complete proteome.  
 FT SEQUENCE 245 AA; 28779 MW; 4597C2B8EF1FA973 CRC64;

Query Match 41.1%; Score 46; DB 1; Length 245;  
 Best Local Similarity 45.0%; Pred. No. 3.3;  
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGVQV 20  
 | ||||:|:| | :|  
 Db 129 TSEITINGELVYNRISGNEI 148

RESULT 15  
 ID LECTI\_MOUSE STANDARD: PRT; 301 AA.  
 AC P24721;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ASIALOGLYCOPROTEIN RECEPTOR 2 (HEPATIC LECTIN 2) (MHL-2) (ASGP-R)  
 DE (ASGPR).



```

GN ASGR2 OR ASGR-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RX MEDLINE=91027942; PubMed=2223888;
RA Sanford J.P., Doyle D.;
RT "Mouse asialoglycoprotein receptor cDNA sequence: conservation of
RT receptor genes during mammalian evolution.";
RL Biochim. Biophys. Acta 1087:259-261(1990).
CC -I- FUNCTION: MEDIATES THE ENDOCYTOSIS OF PLASMA GLYCOPROTEINS TO
CC WHICH THE TERMINAL SIALIC ACID RESIDE ON THEIR COMPLEX
CC CARBOHYDRATE MOETIES HAS BEEN REMOVED. THE RECEPTOR RECOGNIZES
CC TERMINAL GALACTOSE AND N-ACETYLGALACTOSAMINE UNITS. AFTER LIGAND
CC BINDING TO THE RECEPTOR, THE RESULTING COMPLEX IS INTERNALIZED AND
CC TRANSPORTED TO A SORTING ORGANELLE, WHERE RECEPTOR AND LIGAND ARE
CC DISASSOCIATED. THE RECEPTOR THEN RETURNS TO THE CELL MEMBRANE
CC SURFACE.
-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
-I- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN HEPATIC PARENCHYMAL
CC CELLS.
CC -I- MISCELLANEOUS: CALCIUM IS REQUIRED FOR LIGAND BINDING.
CC -I- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53042; CA37211.1; -.
DR PIR; S13165; S13165.
DR HSSP; P06734; 1KJE.
DR MGD; MGI:88082; Asgr2.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
KW Lectin; Glycoprotein; Receptor; Endocytosis; Transmembrane;
KW Calcium; Signal-anchor; Phosphorylation.
FT DOMAIN 1 58 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 59 79 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 80 301 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 169 295 C-TYPE LECTIN (LONG FORM).
FT DISULFID 170 181 BY SIMILARITY.
FT DISULFID 198 293 BY SIMILARITY.
FT DISULFID 271 285 BY SIMILARITY.
FT CARBOHYD 97 97 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 301 AA; 34907 MW; 3A29F1AFBA68F298 CRC64;

Query Match 41.1%; Score 46; DB 1; Length 301;
Best Local Similarity 50.0%; Pred. No. 4.1;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 4 RTINGOLLYFRANGVO 19
DB 153 RTIFCOLAYFQSNQTE 168

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using .sw model

Run on: March 27, 2002, 14:26:02 ; Search time 188.53 Seconds  
(without alignments)  
17.069 Million cell updates/sec

Title: US-09-290-049a-3  
Perfect score: 112  
Sequence: 1 TGAARTINGOLLYFRANGOVYKG 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
al number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_17:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_prodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	80.4	1455	2	069382 streptococc
2	90	80.4	1455	2	069388 streptococc
3	90	80.4	1455	2	069397 streptococc
4	87	77.7	1590	2	059983 streptococc
5	86	76.8	1455	2	069391 streptococc
6	82	73.2	1599	2	000599 streptococc
7	80	71.4	1518	2	000600 streptococc
8	78	69.6	1390	2	069385 streptococc
9	76	67.9	1590	2	055263 streptococc
10	75	67.0	1290	2	048756 streptococc
11	75	67.0	1338	2	09WXJ4 streptococc
12	73	65.2	1449	2	055264 streptococc
13	72	64.3	1477	2	09L466 leuconostoc
14	72	64.3	1508	2	052224 leuconostoc
15	72	64.3	1508	2	09EZH5 leuconostoc
16	72	64.3	1577	2	054178 streptococc
17	68	60.7	1449	2	068542 streptococc
18	68	60.7	1449	2	068542 streptococc
19	66	58.9	1577	2	055265 streptococc

20	65.5	58.5	330	2	055228 streptococc
21	65	58.0	2057	2	09ER05 leuconostoc
22	64	57.1	1512	2	09WXJ5 streptococc
23	57	50.9	1527	2	09ZAR4 leuconostoc
24	57	50.9	1575	2	09LCH3 streptococc
25	54	48.2	221	5	044353 caenorhabdi
26	50	44.6	831	2	050076 clostridium
27	47	42.0	126	2	047171 escherichia
28	47	42.0	126	2	047172 escherichia
29	47	42.0	126	2	047173 escherichia
30	47	42.0	255	2	066378 clostridium
31	47	42.0	396	2	09EP32 clostridium
32	47	42.0	781	11	089090 mus musculu
33	47	42.0	784	11	089087 mus musculu
34	46	41.1	619	2	054972 streptococc
35	46	41.1	1501	2	006307 allicyclobac
36	46	41.1	2364	2	046342 caulobacter
37	45	40.2	142	2	09ABT5 caulobacter
38	45	40.2	181	11	09C021 streptomyce
39	45	40.2	249	2	09RK49 clostridium
40	45	40.2	410	2	032351 arabidopsis
41	45	40.2	587	10	09LDB8 arabidopsis
42	45	40.2	630	10	049501 arabidopsis
43	45	40.2	649	2	003706 clostridium
44	44	39.3	181	4	09ULC4 homo sapien
45	44	39.3	181	11	09DB27 mus musculu

## ALIGNMENTS

RESULT 1	069382	PRELIMINARY;	PRT; 1455 AA.
ID	069382;		
AC	01-AUG-1998 (TREMBLrel. 07, Created)		
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	GLUCOSYLTRANSFERASE-ST.		
GN	GTFC.		
OS	Streptococcus mutans.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Streptococcus.		
OX	NCBI_TaxID=1309;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MT8148;		
RX	MEDLINE=98231643; PubMed=9570124;		
RA	Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,		
RA	Kimura S., Hamada S.;		
RT	"Molecular analyses of glucosyltransferase genes among strains of		
RT	Streptococcus mutans.";		
RL	FEMS Microbiol. Lett. 161:331-336(1998).		
DR	EMBL; D88652; BAA26102.1; -		
DR	InterPro: IPR002479; Glyco_hydro_70.		
DR	InterPro: IPR003318; Glyco_hydro_70.		
DR	Pfam: PF01473; CW_binding_1; 9.		
DR	Pfam: PF02324; Glyco_hydro_70; 1.		
KW	transferase.		
SO	SEQUENCE 1455 AA; 162969 MM; 27D4D3A1ECA2939 CRC64;		
Query Match	80.4%;	Score 90;	DB 2; Length 1455;
Best Local Similarity	72.7%;	Pred. No. 9.2e-06;	
Matches 16;	Conservative 3;	Mismatches 3;	Indels 0; Gaps 0;
QY	1 TGAARTINGOLLYFRANGOVYKG 22		
DB	1329 TGAARTINGOLLYFRANGOVYKG 1350		
RESULT 2	069388		

ID 069388 PRELIMINARY; PRT; 1455 AA.  
 AC 069388;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-ST.  
 GN GTF.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4239;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 Kimura S., Hamada S.;  
 PT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 DR EMBL; D88658; BAA26110.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR Pfam: PF01473; CW\_binding\_1; 10.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match 80.4%; Score 90; DB 2; Length 1455;  
 Best Local Similarity 72.7%; Pred. No. 9.2e-06;  
 Matches 16; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 TGARTINGOLLYFRANGVOYKG 22  
 Db 1329 TGARVNGORLYFKSNGVQAKG 1350

RESULT 3  
 069397 PRELIMINARY; PRT; 1455 AA.  
 AC 069397;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-ST.  
 GN GTF.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4467;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 Kimura S., Hamada S.;  
 PT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 DR EMBL; D89978; BAA26120.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR Pfam: PF01473; CW\_binding\_1; 9.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1455 AA; 162913 MW; A1263427BF24EBE1 CRC64;

Query Match 80.4%; Score 90; DB 2; Length 1455;  
 Best Local Similarity 72.7%; Pred. No. 9.2e-06;  
 Matches 16; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGVOYKG 22  
 Db 1329 TGARVNGORLYFKSNGVQAKG 1350  
 RESULT 4  
 059983 PRELIMINARY; PRT; 1590 AA.  
 ID 059983;  
 AC 059983;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE  
 6-GLUCOSYLTRANSFERASE).  
 GN GTFI.  
 OS Streptococcus sobrinus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1310;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OM2176;  
 RX MEDLINE=94146405; PubMed=8312602;  
 RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.,  
 RT "DNA sequence of the glucosyltransferase gene of serotype d  
 Streptococcus sobrinus.";  
 RL DNA Seq. 4:19-27(1993).  
 CC -I- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) - D-  
 FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
 DR EMBL; D13858; BAA02976.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR Pfam: PF01473; CW\_binding\_1; 16.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Signal; Transferase; Glucosyltransferase.  
 FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1590  
 SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match 77.7%; Score 87; DB 2; Length 1590;  
 Best Local Similarity 77.3%; Pred. No. 3.1e-05;  
 Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 TGARTINGOLLYFRANGVOYKG 22  
 Db 1296 TGAQTIRGOKLYFRKANGVOYKG 1317

RESULT 5  
 069391 PRELIMINARY; PRT; 1455 AA.  
 ID 069391;  
 AC 069391;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-ST.  
 GN GTF.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4251;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 Kimura S., Hamada S.;  
 PT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 DR EMBL; D88661; BAA26114.1; -  
 DR InterPro: IPR002479; CW\_binding.

OS Streptococcus mutans

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4245;  
RX MEDLINE-98231643; PubMed-9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT Streptococcus mutans.";  
RL EMBL: D88655; BAA26106.1; -;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 7.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW transferase.  
SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;

Query Match 69.6%; Score 78; DB 2; Length 1390;  
Best Local Similarity 68.2%; Pred. No. 0.0075;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGARTINGOLTYFRANGVOVGK 22  
DB 1264 TGTVTNGORLXFKRNGVOAKG 1285

RESULT 9  
ID 055263 PRELIMINARY; PRT; 1590 AA.  
AC 055263;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE GTF-I.  
OS Streptococcus sobrinus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1310;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 33478;  
RA Sato S.;  
Ann. Kagoshima Univ. Dental School 16:23-29(1996).  
EMBL: D63570; BAA09792.1; -;  
InterPro: IPR002479; CW\_binding.  
InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 15.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW transferase  
SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 67.9%; Score 76; DB 2; Length 1590;  
Best Local Similarity 68.2%; Pred. No. 0.0018;  
Matches 15; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGARTINGOLTYFRANGVOVGK 22  
DB 1231 TGAQTGKOKLYFEANGGEQVGK 1252

RESULT 10  
ID 048756 PRELIMINARY; PRT; 1290 AA.  
AC 048756;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE DEXTRANSUCRASE.  
"

OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NRRL B1299;  
RX MEDLINE-97136686; PubMed-8982063;  
RA Monchois V., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;  
RT "Cloning and sequencing of a gene coding for a novel dextranucrase  
RT from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-  
RT 6) and alpha (1-3) linkages.";  
RL Gene 182:23-32(1996).  
DR EMBL: U38181; AAB40875.1; -;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 11.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW  
SEQUENCE 1290 AA; 145590 MW; 3555C2E96B749FMA CRC64;

Query Match 67.0%; Score 75; DB 2; Length 1290;  
Best Local Similarity 68.2%; Pred. No. 0.0021;  
Matches 15; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGARTINGOLTYFRANGVOVGK 22  
DB 1245 TGVQVINGOTLTFDADGROYKG 1266

RESULT 11  
ID 09WXJ4 PRELIMINARY; PRT; 1338 AA.  
AC 09WXJ4;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GTF-S.  
GN GTF-S.  
OS Streptococcus criceti.  
OG Plasmid pAM1.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1333;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HS-6;  
RA Inoue M., Fukui K., Miyagi A.;  
RT "S.cricetus glucosyltransferase(gtfs and gtfS) genes.";  
RT Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB026123; BAA77236.1; -;  
DR HSSP: P06278; 1VJ5.  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 10.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Plasmid.  
SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 67.0%; Score 75; DB 2; Length 1338;  
Best Local Similarity 63.6%; Pred. No. 0.0022;  
Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGARTINGOLTYFRANGVOVGK 22  
DB 1207 TGSQTAGOKVFEFOPNGVOVGK 1228

RESULT 12  
ID 055264 PRELIMINARY; PRT; 1449 AA.  
AC 055264;  
OX 055264;

DT		01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE	GLUCOSYLTRANSFERASE PRECURSOR.	
GN	GTF.	
OS	Streptococcus salivarius.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	
OX	NCB1_TaxID=1304;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=95122197; PubMed=78202030;	
RA	Simpson C.L., Giffard P.M., Jacques N.A.;	
RT	"Streptococcus salivarius ATCC 25975 possesses at least two genes	
RL	coding for primer-independent glucosyltransferases.";	
DR	Infect. Immun. 63:609-621(1995).	
DR	EMBL; L35495; AAC41412.1; "	
DR	InterPro; IPR002479; CW_binding.	
DR	InterPro; IPR003318; Glyco_hydro_70.	
DR	Pfam; PF01473; CW_binding_1; 8.	
DR	Pfam; PF02324; Glyco_hydro_70; 1.	
FT	Signal; Transferase.	
FT	SIGNAL	1 35 POTENTIAL.
SO	CHAIN	36 1449 GLUCOSYLTRANSFERASE.
	SEQUENCE	1449 AA; 159984 MW; DD62F07306E86A46 CRC64;
Query Match	Best Local Similarity	65.2%; Score 73; DB 2; Length 1449;
Matches	15; Conservative	2; Mismatches 5; Indels 0; Gaps 0;

RESULT	14	
052224		
ID	052224	PRELIMINARY; PRT: 1508 AA.
AC	052224;	
DT	01-JUN-1998 (TREMblrel. 06, Created)	
DT	01-JUN-1998 (TREMblrel. 06, last sequence update)	
DT	01-JUN-2001 (TREMblrel. 17, last annotation update)	
DE	GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-DE	
GN	DSRB.	
OS	Leucosinocot mesenteroides.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae.	
OX	NCBI_TaxID=1245;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=NRRL B-1299;	
RA	Monchois V., Remaud-Simeon M., Monsan P., Willemot R.M.;	
RL	FEMS Microbiol. Lett. 0:0-0(1998).	
CC	-1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).	
DR	EMBL; AF030129; AAB95453.1; -	
DR	InterPro; IPR002479; CW_binding.	
DR	InterPro; IPR003318; Glyco_hydro_70.	
DR	Pfam; PF01473; CW_binding_1; 14.	
DR	Pfam; PF02324; Glyco_hydro_70; 1.	
KW	Transferase; Glycosyltransferase.	
SO	SEQUENCE 1508 AA; 168511 MW; E70CECB57A/ODI.F0 CRC64;	

RESULT	13		
ID	09L466	PRELIMINARY;	PRT: 1477 AA.
AC	09L466;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	DEXTRANSUCRASE (EC 2.4.1.5).		
GN	DSRC.		
OS	Leuconostoc mesenteroides.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;		
CC	Leuconostoc.		
CX	NCBI_TaxID=1245;		
	[1]		
	SEQUENCE FROM N.A.		
RA	STRAIN-NRRL B-1355;		
RA	Arquellu-Morales M.A., Remaud-Simeon M., Pizut S., Sarcabal P.,		
RT	Willenot R.M., Monsan P.;		
RT	"Sequence analysis of the gene encoding altermansucrase, a sucrose		
RT	glycosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";		
RL	Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.		
DR	EMBL, AJ250172; CAB76565.1; -		
DR	InterPro: IPR002479; CW_binding.		
DR	InterPro: IPR003318; Glyco_hydro_70.		
DR	Pfam: PF01473; CW_binding_1; 14.		
DR	Pfam: PF02324; Glyco_hydro_70; 1.		
KW	transferase; Glycosyltransferase.		
SO	SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFCB831 CRC64;		

```

QY      1 TGARTINGCOLYERANGVOYKG 22
        || ||| | ||| ||| |||
Db      1399 TGLVNINGMLKTYOANGROVKG 1420

RESULT  15

O9EZHS
ID      O9EZHS      PRELIMINARY;      PRT: 1508 AA.
AC      O9EZHS;
DT      01-MAR-2001 (EMBLrel. 16, Created)
DT      01-MAR-2001 (EMBLrel. 16, Last sequence update)
DT      01-JUN-2001 (EMBLrel. 17, Last annotation update)
DE      DEXTRANSUCRase; DSRB742.
GN      DSRB742.
OS      Leuconostoc mesenteroides.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC      Leuconostoc.
OX      NCBI_TaxID=1245;
        [1]
RN      SEQUENCE FROM N.A.
RP      STRAIN=B-742CB,
RA      Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;
        "Leuconostoc mesenteroides B-742CB, a dextranucrase gene.";
RL      Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF294469; AAC38021.1;
DR      InterPro; IPR002479; CW_binding.
DR      InterPro; IPR003118; Glyco_hydro_70.
DR      Pfam; PF01473; CW_binding_1; 14.
DR      Pfam; PF02324; Glyco_hydro_70; 1.
SQ      SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

```

```

Query Match          64.3%   Score 72: DB 2: Length 1477;
Best Local Similarity 68.2%   Pred. No. 0.0075;
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0.

QY      1 TGARTINGCOLLYFRANGVOYKG 22      .
          ||| ||| ||| ||| |||
DB      1368 TGLVINGNGLKYYQANGROYKG 1389

```

Query Match	64.3%	Score 72	DB 2	Length 1508
Best Local Similarity	68.2%	Pred. No. 0.0077		
Matches 15, Conservative	1	Mismatches 6	Indels 0	Gaps 0
QY 1 TGARTINGCOLTFRRANGVQVKG 22				

DB 1399 TGLVNINGNLKRYEQANGROYKG 1420

Search completed: March 27, 2002, 14:26:03  
Job time: 1675 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:49:12 ; Search time 198.55 Seconds

(without alignments)  
7.834 Million cell updates/sec

Title: US-09-290-049a-1

Perfect score: 119  
Sequence: 1 ANDHLSLEAMSDNDPFLYHD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_1101.\*  
1: /SIDSB/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SIDSB/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SIDSB/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SIDSB/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SIDSB/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SIDSB/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SIDSB/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SIDSB/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SIDSB/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SIDSB/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SIDSB/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SIDSB/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SIDSB/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SIDSB/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SIDSB/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SIDSB/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SIDSB/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SIDSB/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SIDSB/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SIDSB/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SIDSB/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
22: /SIDSB/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	90.8	1592	14	AAK32925
2	63	52.9	1577	17	AAK91047
3	60	50.4	2057	21	AAAB1067
4	51	42.9	486	22	AAAB36552
5	50	42.0	305	21	AAAG09845
6	50	42.0	339	21	AAAG09844
7	48.5	40.8	401	22	AAAM43544
8	47.5	39.9	534	21	AAV67414
9	47	39.5	195	21	AAAG09941
10	47	39.5	338	15	AAAR65965
11	46	38.7	118	20	AAV60021

12	46	38.7	199	22	AAAG90817	C glutamicum prote
13	46	38.7	336	20	AAV35244	C. pneumoniae prot
14	45	37.8	236	20	AAV23993	Amino acid sequenc
15	45	37.8	385	12	AAAR15238	Fusarium oxysporum
16	44	37.0	62	21	AAAG44863	Zea mays protein f
17	44	37.0	161	21	AAAG28444	Arabidopsis thalia
18	44	37.0	206	21	AAAG28443	Arabidopsis thalia
19	44	37.0	220	22	AAAG82161	S. epidermidis ope
20	44	37.0	384	21	AAAB29041	C. militans novel
21	43	36.1	126	21	AAAG26416	Arabidopsis thalia
22	43	36.1	126	21	AAAG3894	Arabidopsis thalia
23	43	36.1	170	21	AAAG26415	Arabidopsis thalia
24	43	36.1	170	21	AAAG53893	Arabidopsis thalia
25	43	36.1	193	21	AAAG53892	Arabidopsis thalia
26	43	36.1	194	21	AAAG26414	Arabidopsis thalia
27	43	36.1	327	22	AAAG64817	Human G3PDH fragme
28	43	36.1	335	20	AAV07036	Breast cancer asso
29	43	36.1	335	20	AAV05368	T. thermophilus nit
30	43	36.1	1195	20	AAW75420	Human protein tyro
31	43	36.1	2466	16	AAAR71498	Intracellular prot
32	43	36.1	2466	19	AAW75999	Human prp1 phosph
33	43	36.1	2466	21	AAV90272	Amino acid sequenc
34	43	36.1	2485	22	AAAB19343	Human polypeptide
35	42.5	35.7	325	22	AAAM41942	Corynebacterium gl
36	42.5	35.7	390	22	AAW40156	Arabidopsis thalia
37	42	35.3	123	22	AAAB79865	Human polypeptide
38	42	35.3	174	21	AAAG60587	Human Fc gamma rec
39	42	35.3	178	10	AAAP90061	Arabidopsis thalia
40	42	35.3	185	21	AAAG60586	Arabidopsis thalia
41	42	35.3	189	21	AAAG60585	Expressed antigen
42	42	35.3	236	20	AAW90021	Antigen 1 from glu
43	42	35.3	239	20	AAW89935	Periplasmic chapar
44	42	35.3	259	17	AAAR89326	Helicobacter polyp
45	42	35.3	298	19	AAW71486	

#### ALIGNMENTS

RESULT 1  
ID AAK32925  
AAK32925 standard; Protein; 1592 AA.  
XX  
AC AAK32925;  
XX  
DT 28-JUN-1993 (first entry)  
XX  
DE Glucosyltransferase I.  
XX  
KW GT-1; Streptococcus; dental; carries.  
XX  
OS Streptococcus sobrinus.  
XX  
PN JP05023188-A.  
XX  
XX  
PD 02-FEB-1993.  
XX  
PF 25-JUL-1991; 91JP-0186592.  
XX  
PR 25-JUL-1991; 91JP-0186592.  
XX  
PA (FUKU/) FUKU I.  
PA (KATO/) KATO K.  
XX  
DR WPI; 1993-079449/10.  
XX  
N-PSDB; AAQ37760.  
XX  
PT DNA sequence glucosyl:transferase-I - comprises Streptococcus  
PT sobrinus DNA sequence with at least one nucleotide added or  
XX deleted  
XX  
PS Claim 13; Page 15; 29pp; Japanese.





[illegible]

PR	20-AUG-1999	9905-0149902
PR	23-AUG-1999	9905-0149920
PR	22-AUG-1999	9905-0149930
PR	25-AUG-1999	9905-0150566
PR	26-AUG-1999	9905-0150084
PR	27-AUG-1999	9905-0151065
PR	27-AUG-1999	9905-0151066
PR	27-AUG-1999	9905-0151080
PR	30-AUG-1999	9905-0151303
PR	31-AUG-1999	9905-0151338
PR	01-SEP-1999	9905-0151330
PR	07-SEP-1999	9905-0152363
PR	10-SEP-1999	9905-0153070
PR	13-SEP-1999	9905-0153758
PR	15-SEP-1999	9905-0154018
PR	16-SEP-1999	9905-0154039
PR	20-SEP-1999	9905-0154779
PR	22-SEP-1999	9905-0155139
PR	23-SEP-1999	9905-0155186
PR	24-SEP-1999	9905-0155659
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156596
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158362
PR	13-OCT-1999	9905-0159293
PR	13-OCT-1999	9905-0159294
PR	13-OCT-1999	9905-0159285
PR	14-OCT-1999	9905-0159329
PR	14-OCT-1999	9905-0159330
PR	14-OCT-1999	9905-0159537
PR	14-OCT-1999	9905-0159631
PR	14-OCT-1999	9905-0159638
PR	18-OCT-1999	9905-0159584
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160767
PR	21-OCT-1999	9905-0160768
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160814
PR	21-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160980
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160989
PR	23-OCT-1999	9905-0161404
PR	23-OCT-1999	9905-0161405
PR	23-OCT-1999	9905-0161406
PR	26-OCT-1999	9905-0161359
PR	26-OCT-1999	9905-0161360
PR	26-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161920
PR	28-OCT-1999	9905-0161992
PR	28-OCT-1999	9905-0161993
PR	29-OCT-1999	9905-0162142

Query Match	42.08;	Score 50;	DB 21;	Length 305;
Best Local Similarity	52.98;	Pred. No. 10;		
Matches	9;	Conservative	3;	Mismatches 5;
				Indels 0;
				Gaps 0;

```
QY      3  DHSILEAWSNDNTPYL 19
          | : | : | | | | | :
Db      200 dyegimeawsdkgtllyv 216
```

## RESULT 6

ID AAG09844 standard; Protein; 339 AA.

AC AAG09844;

DT 17-OCT-2000 (first entry)

```
XX Arabidopsis thaliana protein fragment SEQ ID NO: 7934.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EPI033405-A2.
XX
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
```

```
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151330.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      42.0%; Score 50; DB 21; Length 339;
Best Local Similarity 52.9%; Pred. No. 12;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 3 DHSLEAFMSDNDPYL 19
   |:||||| |:
Db 234 dyegimeawskdktly 250

RESULT 7
ID AAM43544
XX AAM43544 standard; Protein; 401 AA.
AC AAM43544:
XX
DE 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 222.
XX
KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; vituicide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
```

```
KW antiinflammatory; antitumor; anticonvulsant; antibacterial;
KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human.
OS Homo sapiens.
XX
PN WO200155308-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01309.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220964.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232168.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
```

PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249254.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249267.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249360.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-488781/53.  
 DR N-PSDB; AAI63850.  
 XX  
 FT New isolated nucleic acids and polypeptides, useful for diagnosing,  
 PT treating and/or preventing human diseases and disorders -  
 XX  
 PS Claim 11; SEQ ID NO 222; 664pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI63803-AAI64012) and  
 CC the encoded proteins (AAM34497-AAM3660) useful for preventing, treating  
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The  
 CC genes were isolated from a range of human tissues disclosed in the  
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists  
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,  
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,  
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 401 AA;  
 Query Match 40.8%; Score 48.5; DB 22; Length 401;  
 Best Local Similarity 47.4%; Pred. No. 24;  
 Matches 9; Conservative 3; Mismatches 4; Indels 3; Gaps 1;  
 QY 5 LSIEMWSDN--DTPYLH 20  
 Db 46 valleywkdntddpxh 64  
 ::||| ||| ||| |||  
 RESULT 8  
 ID AAY67414 standard; protein; 534 AA.  
 XX  
 AC AAY67414;  
 XX  
 DT 12-MAY-2000 (first entry)  
 XX  
 DE Arabidopsis aldehyde dehydrogenase (ALDH)-3.  
 XX  
 KW Arabidopsis; plant plastid; acetyl CoA synthetase; E3 subunit; ACS;  
 KW plant plastidic pyruvate dehydrogenase; PPDH; ATP citrate lyase; ACL;  
 KW pyruvate decarboxylase; PDC; aldehyde dehydrogenase; ALDH; acetyl CoA;  
 KW fatty acid; flavonoid; enzyme; phytochemical; pyruvate decarboxylase;  
 KW acetyl CoA hydrolase; mitochondrial pyruvate dehydrogenase.  
 XX  
 OS Arabidopsis sp.  
 XX  
 PN WO200000619-A2.  
 PN  
 PD 06-JAN-2000.  
 XX  
 XX 25-JUN-1999; 99WO-US14382.  
 PF  
 XX 26-JUN-1998; 98US-0090717.  
 PR  
 XX (IOWA ) UNIV IOWA STATE RES FOUND INC.

XX Nikolau BJ, Murtelle ES, Oliver DJ, Behal R, Schnable PS, Ke J;  
PI Johnson JL, Allred CC, Fattland B, Lutziger I, Wen T;  
XX WPI: 2000-160678/14.  
DR N-PSDB: AA256975.  
XX Novel acetyl CoA synthetase (ACS), plastidic pyruvate dehydrogenase  
PT (PPDH), ATP citrate lyase (ACL), pyruvate decarboxylase (PDC) and  
PT aldehyde dehydrogenase (ALDH) polypeptides used to alter acetyl CoA  
XX levels in plants  
XX Examples: Fig 12B: 79pp: English.  
XX The invention provides nucleic acids encoding Arabidopsis plastidic  
CC acetyl CoA synthetase (ACS), various subunits (specifically the E3  
CC subunit) of plant plastidic pyruvate dehydrogenase (PPDH), the A and B  
CC subunits of a plant ATP citrate lyase (ACL), Arabidopsis pyruvate  
CC decarboxylase (PDC), Arabidopsis aldehyde dehydrogenase (ALDH),  
CC specifically ALDH-2 and ALDH-4. The polypeptides can be expressed by  
CC standard recombinant methodology. The ACS, PPDH, ACL, PDC and ALDH  
CC polypeptides, methods and nucleic acid molecules of the invention are  
CC used to alter the level of acetyl CoA in a plant or plant cell, tissues  
CC or organs. A decrease in acetyl CoA is expected to affect the  
CC biosynthesis of very long chain fatty acids and flavonoids. The enzymes  
CC may also be used for in vitro synthesis of acetyl CoA, which in turn can  
CC be used to produce acetyl CoA phytochemicals (plastidic ACS, PPDH, ACL,  
CC pyruvate decarboxylase, acetyl CoA hydrolase, mitochondrial pyruvate  
CC dehydrogenase and aldehyde dehydrogenase are involved in the biosynthesis  
CC of acetyl CoA.  
XX  
XX Sequence 534 AA:  
SQ  
Query Match 39.9%; Score 47.5; DB 21; Length 534;  
Best Local Similarity 52.9%; Pred. NO. 47;  
Matches 9; Conservative 2; Mismatches 5; Indels 1; Gaps 1;  
QY 2 NDHLSLEMSNDTPY 18  
||: || || || ||  
Db 132 ndelaaletw-dngkpy 147  
RESULT 9  
AAG09941  
ID AAG09941 standard: Protein; 195 AA.  
XX AAG09941:  
XX 17-OCT-2000 (first entry)  
DT  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 8067.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hydrolisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX Arabidopsis thaliana.  
OS  
XX Arabidopsis thaliana.  
PN  
XX EP1033405-A2.  
PD  
XX 06-SEP-2000.  
PF  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 9905-0121825.  
PR 05-MAR-1999; 9905-0123180.  
PR 09-MAR-1999; 9905-0123548.  
PR 23-MAR-1999; 9905-0125788.  
PR 25-MAR-1999; 9905-0126264.  
PR 29-MAR-1999; 9905-0126785.  
PR 01-APR-1999; 9905-0127462.  
PR 06-APR-1999; 9905-0128234.

PR 08-APR-1999; 9905-0128714.  
PR 16-APR-1999; 9905-0129845.  
PR 19-APR-1999; 9905-0130077.  
PR 21-APR-1999; 9905-0130449.  
PR 23-APR-1999; 9905-0130510.  
PR 23-APR-1999; 9905-0130891.  
PR 28-APR-1999; 9905-0131449.  
PR 30-APR-1999; 9905-0132048.  
PR 30-APR-1999; 9905-0132407.  
PR 04-MAY-1999; 9905-0132484.  
PR 05-MAY-1999; 9905-0132485.  
PR 06-MAY-1999; 9905-0132486.  
PR 06-MAY-1999; 9905-0132487.  
PR 07-MAY-1999; 9905-0132863.  
PR 11-MAY-1999; 9905-0134256.  
PR 14-MAY-1999; 9905-0134218.  
PR 14-MAY-1999; 9905-0134219.  
PR 14-MAY-1999; 9905-0134321.  
PR 14-MAY-1999; 9905-0134370.  
PR 18-MAY-1999; 9905-0134768.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135124.  
PR 21-MAY-1999; 9905-0135353.  
PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136392.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139492.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139763.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139899.  
PR 23-JUN-1999; 9905-0140353.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 24-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144088.  
PR 16-JUL-1999; 9905-0144088.  
PR 16-JUL-1999; 9905-0144088.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.



PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144881.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145226.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149920.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 39.5%; Score 47; DB 21; Length 195;  
Best Local Similarity 36.8%; Pred. No. 17;  
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 3 DRLSTLAWSDNDPXYLMD 21

Db 109 nhgevidawshqkplwtd 127

RESULT 10

AAR65965 AAR65965 standard; Protein; 338 AA.

XX AAR65965;

XX 06-JUN-1995 (first entry)

DE T. niveum GAPDH.

XX Alanine-racemase; glyceraldehyde-3-phosphate-dehydrogenase;

KW GAPDH; cyclosporin; immunosuppressive.

XX Tolypocladiinm niveum ATCC 34921.

XX MO9425606-1.

PN 10-NOV-1994.

PD 23-APR-1994; 94WO-EP01272.

XX 23-APR-1993; 93DE-4312856.

PR 04-MAY-1993; 93DE-4314610.

PR 17-MAY-1993; 93DE-4316419.

XX (SANO ) SANDOZ LTD.

PA (SANO ) SANDOZ PATENT GMBH.

PA (SANO ) SANDOZ-ERFINDUNGEN VERW GES MBH.

PI Koehrer HP, Schneider-scherzer E, Schoergendorfer K;

XX Weber G;

XX WPI: 1994-358281/44.

New nucleic acid encoding eukaryotic alanine racemase - and  
PT related vectors, host cells and recombinant enzyme, useful for  
PT producing cyclosporin derivs. or increasing cyclosporin prodn.,



XX  
SQ Sequence 199 AA;

Query Match 38.7%; Score 46; DB 22; Length 199;  
Best Local Similarity 41.2%; Pred. No. 25;  
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 NDHLSILEAMSDNDPPY 18  
:|::|:|:|:|  
Db 172 ddsqdlvewldqdlrly 188

RESULT 13  
AAI35244  
ID AAI35244 standard; Protein: 336 AA.

XX  
AC AAI35244;

XX  
DT 13-SEP-1999 (first entry)

XX C. pneumoniae protein involved in intermediate metabolism.

KM Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN WO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-1B01890.

XX 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GEST ) GENSET.

PI Griffiths R;

DR WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

Page 1086-1087; Disclosure: 1912pp; English.  
AAI34584-Y35879 represent the proteins encoded by all the open reading  
frames in the complete genome (see AAI34584) of Chlamydia pneumoniae.

C. pneumoniae causes respiratory disease such as pneumonia and  
bronchitis and is thought to be a contributing factor in heart  
disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
nodosum or pharyngitis. The polypeptides encoded by the open reading  
frames of the C. pneumoniae genome (see AAI34584-Y35879) can be used in  
immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
nucleotide sequences can also be used as immunogenic compositions,  
especially where the vector directs the expression of a neutralising  
epitope of C. pneumoniae.

XX  
SQ Sequence 336 AA;

Query Match 38.7%; Score 46; DB 20; Length 336;  
Best Local Similarity 52.9%; Pred. No. 46;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 2 NDHLSILEAMSDNDPPY 18  
|||:|:|:|:|  
Db 304 ndrftklvawydnctgy 320

RESULT 14

AAI23993  
ID AAI23993 standard; Protein: 236 AA.

XX  
AC AAI23993;

DT 22-SEP-1999 (first entry)

DE Amino acid sequence of a human PIG protein.

KW p53 transcription tag; p53 status; cancer; cytotoxicity; carcinogenicity;  
KW neoplastic; PIG.

OS Homo sapiens.

PN WO9914356-A2.

PD 25-MAR-1999.

PF 17-SEP-1998; 98WO-US19300.

PR 30-MAR-1998; 98US-0079817.

PR 17-SEP-1997; 97US-0059153.

XX (UYJO ) UNIV JOHNS HOPKINS.

PI Kinzler KW, Polyak K, Vogelstein B;

DR WPI; 1999-443793/37.

XX Use of p53 transcription tags to determine p53 status in, e.g.  
cancer diagnosis

PS Claim 39; Page 67; 73pp; English.

The specification describes the use of p53 transcription tags for  
developing products to determine p53 status, to diagnose cancer  
and to evaluate cytotoxicity or carcinogenicity of a test agent.  
A method for diagnosing cancer or determining p53 status in a sample  
suspected for being neoplastic comprises comparing the level of  
transcription of an RNA transcript in a first sample (s1) of a first  
tissue (t1) to the level of transcription of the transcript in a second  
sample (s2) of a second tissue (s2), where s1 is suspected of being  
neoplastic and s2 is a normal human tissue (of the same type) and the  
transcript is identified by a tag, and categorizing s1 as neoplastic  
or as having a mutant p53 when transcription is found to be the same  
or lower in the first, than in s2. The methods and products can be used  
to determine p53 status, to diagnose cancer and to evaluate cytotoxicity  
or carcinogenicity of a test agent. AAI23992-Y24000 represent human PIG  
proteins.

XX  
SQ Sequence 236 AA;

Query Match 37.8%; Score 45; DB 20; Length 236;  
Best Local Similarity 44.0%; Pred. No. 44;  
Matches 11; Conservative 3; Mismatches 3; Indels 8; Gaps 2;

OY 1 ANDHLSIL-----EAMSDNDPPYLH 20  
||:|:|:|:|:|:|:|:|  
Db 30 adthrlrlgkpwmerw---etpymh 51

RESULT 15

AAI15238  
ID AAI15238 standard; Protein: 385 AA.

XX  
AC AAI15238;

DT 11-FEB-1992 (first entry)

XX Fusarium oxysporum F-family cellobiohydrolase.

XX cellulose; carbohydrate binding domain; CBD; fungi; cellulase.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:38 ; Search time 53.4 Seconds

(without alignments)  
14.419 Million cell updates/sec

Title: US-09-290-049a-1  
Perfect score: 119  
Sequence: 1 ANDHLSTLEAMSNDPRYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_39:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	1476	1	GTFB_STRMU
2	110	92.4	1375	1	GTFB_STRMU
3	108	90.8	1592	1	GTFB_STRMU
4	108	90.8	1597	1	GTFB_STRMU
5	78	65.5	1365	1	GTFB_STRMU
6	75	63.0	1462	1	GTFB_STRMU
7	57	47.9	330	1	GTFB_STRMU
8	49	41.2	336	1	GTFB_STRMU
9	47.5	39.9	330	1	GTFB_STRMU
10	46	38.7	275	1	GTFB_STRMU
11	46	38.7	335	1	GTFB_STRMU
12	46	38.7	476	1	GTFB_STRMU
13	45	37.8	236	1	GTFB_STRMU
14	45	37.8	385	1	GTFB_STRMU
15	44.5	37.4	470	1	GTFB_STRMU
16	44	37.0	234	1	GTFB_STRMU
17	44	37.0	330	1	GTFB_STRMU
18	44	37.0	332	1	GTFB_STRMU
19	44	37.0	332	1	GTFB_STRMU
20	44	37.0	337	1	GTFB_STRMU
21	44	37.0	337	1	GTFB_STRMU
22	44	37.0	337	1	GTFB_STRMU
23	44	37.0	337	1	GTFB_STRMU
24	44	37.0	337	1	GTFB_STRMU
25	44	37.0	337	1	GTFB_STRMU
26	43.5	36.6	115	1	GTFB_STRMU
27	43.5	36.6	130	1	GTFB_STRMU
28	43.5	36.6	130	1	GTFB_STRMU
29	43.5	36.6	130	1	GTFB_STRMU
30	43.5	36.6	130	1	GTFB_STRMU
31	43.5	36.6	130	1	GTFB_STRMU
32	43.5	36.6	130	1	GTFB_STRMU
33	43.5	36.6	130	1	GTFB_STRMU

34	43	36.1	322	1	G3P_SHEEP
35	43	36.1	332	1	G3P_PIG
36	43	36.1	332	1	G3P_RABIT
37	43	36.1	334	1	G3P_HUMAN
38	43	36.1	337	1	G3P_MYGE
39	43	36.1	337	1	G3P_MYGE
40	43	36.1	424	1	YD_RAT
41	43	36.1	674	1	CONA_BACTJ
42	42.5	35.7	2485	1	PTND_HUMAN
43	42.5	35.7	441	1	HGD_STRCO
44	42.5	35.7	491	1	GPD_ECOLI
45	42.5	35.7	645	1	IME2_YEAST
			764	1	Y157_YEAST

## ALIGNMENTS

RESULT 1  
ID GTFB\_STRMU STANDARD: PRT: 1476 AA.  
AC P08987; 069381; 069384; 069387; 069390; 069396;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLUCOSYLTRANSFERASE-T PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
DE (SUEROSE 6-GLUCOSYLTRANSFERASE).  
GN GTFB.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OX Streptococcus.  
NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GS-5;  
RX MEDLINE=87308013; PubMed=3040685;  
RA Shiroza T., Ueda S., Kuramitsu H.K.;  
RT "Sequence analysis of the gtfb gene from Streptococcus mutans.";  
RL J. Bacteriol. 169:4263-4270(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -I- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) - D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).  
CC -I- SUBCELLULAR LOCATION: SECRETED.  
CC -I- DISEASE: DENTAL CARIES.  
CC -I- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.  
CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S. MUTANS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: M17361; AAA8588.1; -  
CC EMBL: D88651; BAA26101.1; -  
CC EMBL: D88654; BAA26105.1; -

DR EMBL: D8657; BAA26109.1; -  
 DR EMBL: D8660; BAA26113.1; -  
 DR EMBL: D89977; BAA26119.1; -  
 DR PIR: B3135; B3135.  
 DR InterPro: IPR002479; CW\_binding\_70.  
 DR Pfam: PF01473; Glyco\_hydro\_70; 1.  
 DR Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 KW SIGNAL  
 FT CHAIN 1 34  
 FT SIGNAL 35 1476  
 FT DOMAIN 35 1051  
 FT REPEAT 1097 1476  
 FT REPEAT 1097 1476  
 FT REPEAT 1161 1470  
 FT REPEAT 1161 1470  
 FT REPEAT 1225 1275  
 FT REPEAT 1290 1340  
 FT REPEAT 1355 1405  
 FT REPEAT 1420 1470  
 FT REPEAT 1420 1470  
 FT VARIANT 62 62  
 FT VARIANT 65 65  
 FT VARIANT 65 65  
 FT VARIANT 68 68  
 FT VARIANT 78 78  
 FT VARIANT 86 86  
 FT VARIANT 89 89  
 FT VARIANT 168 168  
 FT VARIANT 276 276  
 FT VARIANT 399 399  
 FT VARIANT 474 474  
 FT VARIANT 512 512  
 FT VARIANT 519 519  
 FT VARIANT 701 701  
 FT VARIANT 708 708  
 FT VARIANT 938 938  
 FT VARIANT 952 957  
 FT VARIANT 963 964  
 FT VARIANT 968 970  
 FT VARIANT 1086 1086  
 FT VARIANT 1158 1158  
 FT VARIANT 1163 1163  
 FT VARIANT 1168 1168  
 FT VARIANT 1182 1182  
 FT VARIANT 1234 1234  
 FT VARIANT 1263 1263  
 FT VARIANT 1263 1263  
 FT VARIANT 1264 1264  
 FT VARIANT 1272 1272  
 FT VARIANT 1329 1329  
 FT VARIANT 1394 1394  
 FT VARIANT 1402 1402  
 FT VARIANT 1459 1459  
 FT CONFLICT 570 570  
 FT CONFLICT 800 817  
 FT CONFLICT 1310 1310  
 FT SEQUENCE 1476 AA; 165688 MW; 3479862807694D98 CRC64;

Query Match 100.0%; Score 119; DB 1; Length 1476;  
 Best Local Similarity 100.0%; Pred. No. 2,1e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHSLTEAMSDNDTPYLHD 21  
 DB 481 ANDHSLTEAMSDNDTPYLHD 501

RESULT 2  
 ID GTF-STRM0 STANDARD; PRT: 1375 AA.  
 AC P13470; P05427;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)  
 DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTF-STRM0  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBL\_TaxID-1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GS-5;  
 RC MEDLINE-89137980; PubMed-2976010;  
 RA Ueda S., Shiroza T., Kuramitsu H.K.;  
 RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";  
 RL gene 69:101-109(1988).  
 RL [2]  
 RP SEQUENCE OF 1-349 FROM N.A.  
 RC STRAIN-GS-5;  
 RC MEDLINE-87308013; PubMed-3040685;  
 RA Shiroza T., Ueda S., Kuramitsu H.K.;  
 RT "Sequence analysis of the gtfB gene from Streptococcus mutans GS-5.";  
 RL J. Bacteriol. 169:4263-4270(1987).  
 RL [1]  
 RP FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -  
 CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 1,3-LINKED GLUCOSE AND SOME 1,6-GLUCOSE). GTF-S SYNTHESIZES BOTH  
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 FORMS OF GLUCANS.  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: M22054; AAA88592.1; -  
 CC EMBL: M17361; AAA88589.1; -  
 DR PIR: J70345; J70345.  
 DR PIR: C3135; C3135.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; Glyco\_hydro\_70; 1.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 34  
 FT SIGNAL 35 1375  
 FT DOMAIN 35 1050  
 FT DOMAIN 1126 1375  
 FT DOMAIN 1126 1375  
 FT REPEAT 1126 1159  
 FT REPEAT 1169 1200  
 FT REPEAT 1169 1200  
 FT REPEAT 1227 1238  
 FT REPEAT 1227 1238  
 FT REPEAT 1253 1303  
 FT REPEAT 1318 1330  
 FT REPEAT 1318 1330  
 FT SEQUENCE 1375 AA; 153022 MW; D4880CB80ACF13 CRC64;

Query Match 92.4%; Score 110; DB 1; Length 1375;  
 Best Local Similarity 95.2%; Pred. No. 4.7e-09;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDPTPLYLD 21  
 |||  
 DB 507 ANDHLSILEAMSYNDPTPLYLD 527

RESULT 3  
 GTF2\_STRDO STANDARD; PRT; 1592 AA.  
 ID GTF2\_STRDO

DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE (GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 Streptococcus downei (Streptococcus sobrinus).  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.  
 NCBI\_TaxID=1317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=6715;  
 RX MEDLINE=9112327; PubMed=1704006;  
 RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,  
 Kagawa H.;  
 RT "Peptide sequences for sucrose splitting and glucan binding within  
 streptococcus sobrinus glucosyltransferase (water-insoluble glucan  
 synthase)." ;  
 RL J. Bacteriol. 173:989-996(1991).  
 CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -I- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -I- SUBCELLULAR LOCATION: SECRETED.  
 CC -I- DISEASE: DENTAL CARIES.  
 CC -I- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 FORMS OF GLUCANS.  
 CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S.MUTANS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: D90213; BAA14241.1; -  
 CC PTR: A38175; A38175.  
 DR HSP; P00695; ZHE.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 16.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-1.  
 FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1093 1592 6.5 X TANDEM REPEATS.  
 FT REPEAT 1093 1142 1.  
 FT REPEAT 1158 1207 2.  
 FT REPEAT 1222 1272 3.

FT REPEAT 1287 1337 4.  
 FT REPEAT 1402 1451 5.  
 FT REPEAT 1514 1563 6.  
 FT REPEAT 1577 1592 7 (INCOMPLETE).  
 SQ SEQUENCE 1592 AA; 176167 MW; BCOA66D079351ECF CRC64;

Query Match 90.8%; Score 108; DB 1; Length 1592;  
 Best Local Similarity 85.7%; Pred. No. 1.1e-08;  
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDPTPLYLD 21  
 |||  
 DB 477 ANNHVSIVEAMSDNDPTPLYLD 497

RESULT 4  
 GTF1\_STRDO STANDARD; PRT; 1597 AA.  
 ID GTF1\_STRDO

DT 01-JUL-1988 (Rel. 11, Created)  
 DT 01-JUL-1988 (Rel. 11, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 Streptococcus downei (Streptococcus sobrinus).  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.  
 NCBI\_TaxID=1317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MFE28;  
 RX MEDLINE=87308014; PubMed=3040686;  
 RA Ferretti V.J., Gilpin M.L., Russell R.R.B.;  
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus  
 sobrinus Mfe28." ;  
 RL J. Bacteriol. 169:4271-4278(1987).  
 CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -I- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -I- SUBCELLULAR LOCATION: SECRETED.  
 CC -I- DISEASE: DENTAL CARIES.  
 CC -I- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 FORMS OF GLUCANS.  
 CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S.MUTANS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M17391; AAC63063.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 19.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1597 GLUCOSYLTRANSFERASE-1.  
 FT DOMAIN 39 1050 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.  
 FT REPEAT 1099 1132 A REPEAT.

FT REPEAT 1163 1213 AC REPEAT.  
 FT REPEAT 1227 1277 AC REPEAT.  
 FT REPEAT 1292 1342 AC REPEAT.  
 FT REPEAT 1352 1399 B REPEAT.  
 FT REPEAT 1406 1455 AC REPEAT.  
 FT REPEAT 1465 1512 B REPEAT.  
 FT REPEAT 1519 1568 AC REPEAT.  
 FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).  
 SQ SEQUENCE 1597 AA; 177080 MW; B9E6A200686798E CRC64;

Query Match 90.8%; Score 108; DB 1; Length 1597;  
 Best Local Similarity 85.7%; Pred. No. 1.2e-08;  
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEMSDNDPPLYLD 21  
 DB 483 ANNHVSIVAWSNDPPLYLD 503

## RESULT 5

GFPS\_STRDO STANDARD; PRT: 1365 AA.  
 ID GFPS\_STRDO  
 AC P29336;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTF-S.  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MFE28;  
 RX MEDLINE-90316665; PubMed-2142479;  
 RA Gilmore K.S., Russell R.R., Ferretti J.J.;  
 RT "Analysis of the Streptococcus downei gtfS gene, which specifies a  
 glucosyltransferase that synthesizes soluble glucans.";  
 RL Infect. Immun. 58:2452-2458(1990).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
 PRIMER GLUCAN UNLIKE GTF-I.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA  
 1,6-GLUCOSE).  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M30943; AAA26898.1; -.  
 DR PIR: A41483; A41483.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 10.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 36  
 FT CHAIN 37 1365 OR 37 (POTENTIAL).  
 FT GLUCOSYLTRANSFERASE-S.

FT DOMAIN 37 1050 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1083 1365 4.5 X TANDEN REPEATS.  
 FT REPEAT 1083 1131 1.  
 FT REPEAT 1150 1199 2.  
 FT REPEAT 1225 1274 3.  
 FT REPEAT 1289 1339 4.  
 FT REPEAT 1353 1365 5 (INCOMPLETE).  
 SQ SEQUENCE 1365 AA; 151590 MW; 167296B5A2E8C476 CRC64;

Query Match 65.5%; Score 78; DB 1; Length 1365;  
 Best Local Similarity 71.4%; Pred. No. 0.00044;  
 Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSILEMSDNDPPLYLD 21  
 DB 467 AIDHLSILEMSGNDNDVYKD 487

## RESULT 6

GFPS\_STRDO STANDARD; PRT: 1462 AA.  
 ID GFPS\_STRDO  
 AC P49331; O69383; O69386; O69389; O69392; O69398;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFD.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GS-5;  
 RX MEDLINE-91100958; PubMed-2148600;  
 RA Honda O., Kato C., Kuramitsu H.K.;  
 RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
 the glucosyltransferase-S enzyme.";  
 RL J. Gen. Microbiol. 136:2099-2105(1990).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN-MT4239, MT4245, MT4251, MT4467, AND MT8148;  
 CC MEDLINE-98231643; PubMed-9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
 RA Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 RL FEMS Microbiol. Lett. 161:331-336(1998).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 FORMS OF GLUCANS.  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----



[illegible]

```

Query Match      63.0%   Score 75;   DB 1;   Length 1462;
Best Local Similarity 71.4%   Pred. No. 0.0014;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0.

OY      1 ANDHLSILEAWMSDNDTPYLLHD 21
      | : ||||| ||||| || ||

```

Db	495	AINHLSILEAMSDNDNPQYNKD	515
RESULT	7		
ID	G3PC_LEIEME	STANDARD;	PRT; 330 AA.
AC	001558:		
DT	01-APR-1993 (Rel. 25, Created)		
DT	01-FEB-1994 (Rel. 28, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (EC 1.2.1.12)		
DE	(GAPDH).		
GN	GAPC.		
OS	Leishmania mexicana.		
OC	Eukaryota, Euklenozoa, Kinetoplastida; Trypanosomatidae; Leishmania.		
OX	NCBI_TaxID=5665;		
RA	SEQUENCE FROM M.X.A.		
RC	STRAIN=SSP. MEXICANA;		
RX	MEDLINE=93063042; PubMed=1435864;		
RA	Hannaert V., Blaauw M., Kohl L., Allert S., Operdoes F.R.,		
RT	Michels P.A.M.;		
RT	"Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-		
RL	phosphate dehydrogenase in Leishmania mexicana.";		
RL	Mol. Biochem. Parasitol. 55:115-126(1992).		
CC	-1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE		
CC	+ NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.		
CC	-1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.		
CC	-1- SUBUNIT: HOMOTETRAMER.		
CC	-1- SUBCELLULAR LOCATION: CYTOSOL.		
CC	-1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE		
CC	DEHYDROGENASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@lsb-sib.ch">license@lsb-sib.ch</a> ).		
CC	-----		
DR	EMBL; X65220; CMAA6323.1; .		
DR	PIR; S25142; S25142.		
DR	PIR; B48445; B48445.		
DR	HSSP; P06977; 1GAD.		
DR	InterPro; IPR000173; GAP_DH.		
DR	Pfam; PF00044; gpdb.1.		
DR	PRINTS; PR00078; G3PDHGNASE.		
DR	PROSITE; PS00071; GAPDH.1.		
FW	GLYCOLYSIS; Oxidoreductase; NAD.		
FT	INIT_MET 0 0		
FT	BY SIMILARITY.		
FT	BINDING 148 148		
FT	ACT_SITE 175 175		
FT	ACTIVATES THIOL GROUP DURING CATALYSIS.		
SO	SEQUENCE 330 AA; 35511 MW; ED4B6D8EA207FLE CRC64;		
QY	2 NDHLSILEAMSDNDNTPYLH 20		
Db	300 NDHLEFKLVSMWDNERTGSH 318		
Query Match	47.9%;	Score 57;	DB 1;
Best Local Similarity	52.6%;	Pred. 0.15;	Length 330;
Matches 10;	Conservative 2;	Mismatches 7;	Indels 0;
Gaps	0;		
RESULT	8		
ID	UL16_EBV	STANDARD;	PRT; 336 AA.
AC	P03221;		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	21-JUL-1986 (Rel. 01, Last sequence update)		
DT	15-FEB-1998 (Rel. 37, Last annotation update)		
DE	BGLF2 PROTEIN.		



```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83099; CAAS8155.1; -.
DR EMBL; Z72905; CA97130.1; -.
DR SGD; S0003352; SEC35.
KW Hypothetical protein.
SO SEQUENCE 275 AA; 31799 MW; AA102D086F3FAD7 CRC64;

Query Match 38.7%, Score 46; DB 1; Length 275;
Best Local Similarity 50.0%; Pred. No. 6.1;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 7 ILEAMSDNDTPYLH 20
   ||| :|::|::|::|
217 ILESCADSNSPYIH 230

RESULT 11
G3P_CHLPHN STANDARD; PRT; 335 AA.
ID G3P_CHLPHN Q9J0H7;
AC Q9J0H7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH).
GN GAP OR GAPA OR CPND624 OR CP0123.
OS Chlamydia pneumoniae (Chlamydochlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydochlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baas S.,
RA Linher M., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Givani M., Neilson W., DeBooy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + OXTHOPHOSPHATE
CC + NAD(+) = 1,3-DIPHOSPHATGLYCERATE + NADH.
CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; AE001647; AAD18763.1; -
CC DR EMBL; AE002173; AAF38006.1; -
CC DR EMBL; AP002547; BAA98831.1; -
CC DR HSSP; P06977; 1GAE.
CC DR TIGR; CP0123; -.
CC DR InterPro; IPR000173; GAP_DH.
CC DR Pfam; PF00044; gpdh; 1.
CC DR PRINTS; PR00078; G3PDHGRGNASE.
CC DR PROSITE; PS00071; GAPDH; 1.
CC DR GlycoLysis; Oxidoreductase; NAD: Complete proteome.
CC FT BINDING 151 151 GLYCERALDEHYDE 3-PHOSPHATE (BY
CC FT ACT_SITE 178 178 SIMILARITY).
CC FT ACT_SITE 178 178 ACTIVATES THIOL GROUP DURING CATALYSIS
CC FT SEQUENCE 335 AA; 36837 MM; C66DDE3AD3ADF8FE CRC64;
CC SQ
CC
CC Query Match 38.7%; Score 46; DB 1; Length 335;
CC Best Local Similarity 52.9%; Pred. No. 7.7;
CC Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
CC
CC QY 2 NDHSLILEAMSGNDPPY 18
CC || || || || || ||
CC Db 303 NDREFKLVAMWDENERGY 319
CC
CC RESULT 12
CC PCNA_PSEAE
CC ID PCNA_PSEAE STANDARD: PRT; 476 AA.
CC AC Q51508.
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 20-NOV-2001 (Rel. 40, Last annotation update)
CC DE SALICYLATE BIOSYNTHESIS ISOCHORISMATE SYNTHASE (EC 5.4.99.6).
CC GN PCNA OR PA4231.
CC OS Pseudomonas aeruginosa.
CC OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
CC OC Pseudomonas.
CC OX NCBI_TaxID=287;
CC
CC [1]
CC RN SEQUENCE FROM N.A.
CC RP STRAIN-ATCC 15692 / PAOI;
CC RC MEDLINE=96086939; PubMed=7500944;
CC RX Serino L., Reilmann C., Baur H., Beyeler M., Visca P., Haas D.;
CC RT "Structural genes for salicylate biosynthesis from chorismate in
CC RL Pseudomonas aeruginosa.";
CC RL Mol. Gen. Genet. 249:217-228(1995).
CC [2]
CC RN SEQUENCE FROM N.A.
CC RP STRAIN-ATCC 15692 / PAOI;
CC RC MEDLINE=20437337; PubMed=10984043;
CC RX Stover C.K., Pham X.-Q.T., Ervin A.L., Mizoguchi S.D., Warren P.,
CC RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
CC RA Gader R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
CC RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Lartig K., Tam R.M.,
CC RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
CC RA Reizer J., Slier M.H., Hancock R.E.W., Lory S., Olson M.V.;
CC RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
CC RT opportunistic pathogen.";
CC RT Nature 406:959-964(2000).
CC CC -1- FUNCTION: INVOLVED IN THE CONVERSION OF CHORISMATE TO SALICYLATE
CC CC (PROBABLY).
CC CC -1- CATALYTIC ACTIVITY: CHORISMATE - ISOCHORISMATE.
CC CC -1- PATHWAY: SALICYLATE BIOSYNTHESIS.
CC CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO
CC TRE AND PABH.

```





...

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:58:08 ; Search time 188.53 Seconds  
(without alignments)  
16.293 Million cell updates/sec

Title: US-09-290-049a-1  
Perfect score: 119  
Sequence: 1 ANDHSLILEAMSDNDPPLYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP.TREMBL\_17:\*  
2: SP.archaea:\*  
3: SP.bacteria:\*  
4: SP.fungi:\*  
5: SP.human:\*  
6: SP.invertebrate:\*  
7: SP.mammal:\*  
8: SP.mhc:\*  
9: SP.organelle:\*  
10: SP.phage:\*  
11: SP.plant:\*  
12: SP.rodent:\*  
13: SP.virus:\*  
14: SP.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	1390	2	069385 streptococc
2	119	100.0	1455	2	069391 streptococc
3	110	92.4	1435	2	069382 streptococc
4	110	92.4	1455	2	069388 streptococc
5	110	92.4	1455	2	069397 streptococc
6	108	90.8	1590	2	055263 streptococc
7	108	90.8	1590	2	059983 streptococc
8	78	65.5	1338	2	09WXRJ4 streptococc
9	76	63.9	1016	2	09LJCJ7 leuconostoc
10	74	62.2	1477	2	09LJ466 leuconostoc
11	74	62.2	1508	2	052224 leuconostoc
12	74	62.2	1508	2	09EZH5 leuconostoc
13	73	61.3	1575	2	09LCH3 streptococc
14	73	61.3	1577	2	054178 streptococc
15	71	59.7	1527	2	09ZARA4 leuconostoc
16	66	55.5	1449	2	068542 streptococc
17	66	55.5	1449	2	055264 streptococc
18	63	52.9	1577	2	055265 streptococc
19	62	52.1	1512	2	09WXRJ5 streptococc

20	60	50.4	2057	2	09RE05 leuconostoc
21	58	48.7	1518	2	000600 streptococc
22	57	47.9	1599	2	000599 streptococc
23	52	43.7	4848	2	007944 streptomyc
24	51	42.9	435	10	09LJ68 arabidopsis
25	50	42.0	347	10	09LY92 arabidopsis
26	49.5	41.6	549	10	043274 zea mays (m
27	49	41.2	601	2	005566 mycobacteri
28	48	40.3	490	2	006579 mycobacteri
29	47.5	39.9	519	10	09ZUB6 arabidopsis
30	47.5	39.9	549	10	09LJL2 oryza sativ
31	47.5	39.9	549	10	09FRX7 oryza sativ
32	47	39.5	175	10	09LJM8 arabidopsis
33	47	39.5	378	10	09SZH3 arabidopsis
34	47	39.5	524	2	09KST8 vibrio chol
35	47	39.5	525	2	039679 schizosacch
36	46	38.7	195	13	09DFM6 gillilichthys
37	46	38.7	2567	5	09U0J6 plasmodium
38	45.5	38.2	96	2	09JST4 neisseria m
39	45	37.8	286	2	09RND2 streptomyc
40	45	37.8	333	13	09PTW5 arabidopsis
41	45	37.8	947	10	023136 homo sapien
42	45	37.8	1159	4	09UFC39 pseudomonas
43	44.5	37.4	210	2	09I0Y4 drosophila
44	44.5	37.4	499	5	09VJ56 homo sapien
45	44	37.0	44	4	016768 homo sapien

#### ALIGNMENTS

RESULT 1  
ID 069385 PRELIMINARY: PRT; 1390 AA.  
AC 069385;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTFC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4245;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.,  
RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL: D88655; BAA26106.1; .  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 7.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SO SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;

Query Match 100.0%; Score 119; DB 2; Length 1390;  
Best Local Similarity 100.0%; Pred. No. 7.5e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHSLILEAMSDNDPPLYLHD 21  
DB 507 ANDHSLILEAMSDNDPPLYLHD 527  
RESULT 2  
069391

ID 069391 PRELIMINARY; PRT: 1455 AA.  
AC 069391;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4251;  
RC MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans.";  
EMBL: D88651; BAA26114.1; -  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 9.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 100.0%; Score 119; DB 2; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 7.9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 ANDHLSILEAMSDNDTPYLHD 21  
DB 507 ANDHLSILEAMSDNDTPYLHD 527

RESULT 3  
ID 069382 PRELIMINARY; PRT: 1455 AA.  
AC 069382;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT8148;  
RC MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans.";  
EMBL: D88652; BAA26102.1; -  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 9.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162969 MW; 27DAD3A1ECCA2939 CRC64;

Query Match 92.4%; Score 110; DB 2; Length 1455;  
Best Local Similarity 95.2%; Pred. No. 2e-08;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDTPYLHD 21  
DB 507 ANDHLSILEAMSYNDTPYLHD 527

RESULT 4  
ID 069388 PRELIMINARY; PRT: 1455 AA.  
AC 069388;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4239;  
RC MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans.";  
EMBL: D88658; BAA26110.1; -  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 10.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match 92.4%; Score 110; DB 2; Length 1455;  
Best Local Similarity 95.2%; Pred. No. 2e-08;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDTPYLHD 21  
DB 507 ANDHLSILEAMSYNDTPYLHD 527

RESULT 5  
ID 069397 PRELIMINARY; PRT: 1455 AA.  
AC 069397;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4467;  
RC MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans.";  
EMBL: D89978; BAA26120.1; -  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 9.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.



SQ SEQUENCE 1455 AA; 162913 MW; A1263427BF24E8E1 CRC64;

Query Match 92.4%; Score 110; DB 2; Length 1455;

Best Local Similarity 95.2%; Pred. No. 2e-08;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHSTILEAWSNDNDPYLHD 21

DB 507 ANDHSTILEAWSNDNDPYLHD 527

RESULT 6

ID 055263 PRELIMINARY; PRT: 1590 AA.

AC 035263;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE GTF-1.

Streptococcus sobrinus.

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

Streptococcus.

NCBI\_TaxID=1310;

OX 11

RN SEQUENCE FROM N.A.

RC STRAIN=ATCC 33478;

RA Sato S.;

RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).

DR EMBL; D63570; BAA09792.1; "

DR Interpro: IPR002479; CW\_binding.

DR Interpro: IPR003318; Glyco\_hydro\_70.

DR Pfam; PF04473; CW\_binding\_1; 15.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Transferase.

SQ SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 90.8%; Score 108; DB 2; Length 1590;

Best Local Similarity 85.7%; Pred. No. 4.5e-08;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHSTILEAWSNDNDPYLHD 21

DB 477 ANNHVSIVEAWSNDNDPYLHD 497

RESULT 7

ID 059983 PRELIMINARY; PRT: 1590 AA.

AC 059983;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE

6-GLUCOSYLTRANSFERASE).

GN GTF1.

OS Streptococcus sobrinus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1310;

RN SEQUENCE FROM N.A.

RC STRAIN=OM2176;

RA MEDLINE=94146405; PubMed=8312602;

RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;

RT "DNA sequence of the glucosyltransferase gene of serotype d

Streptococcus sobrinus.";

RL DNA Seq. 4:19-27(1993).

-i- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-

CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).

DR EMBL; D13858; BAA02976.1; "

DR Interpro: IPR002479; CW\_binding.

DR Interpro: IPR003318; Glyco\_hydro\_70.

DR Pfam; PF04473; CW\_binding\_1; 16.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Signal; Transferase; Glycosyltransferase.

FT SIGNAL 1 38 POTENTIAL.

CHAIN 39 1590 GLUCOSYLTRANSFERASE-1.

SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57C3C2B0E CRC64;

Query Match 90.8%; Score 108; DB 2; Length 1590;

Best Local Similarity 85.7%; Pred. No. 4.5e-08;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHSTILEAWSNDNDPYLHD 21

DB 477 ANNHVSIVEAWSNDNDPYLHD 497

RESULT 8

ID 09MXJ4 PRELIMINARY; PRT: 1338 AA.

AC 09MXJ4;

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GTF-S.

GN Streptococcus criceti.

OS Plasmid pAM1.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

NCBI\_TaxID=1333;

OX 11

RN SEQUENCE FROM N.A.

RC STRAIN=HS-6;

RA Inoue M., Fukui K., Miyagi A.;

RT "S. cricetus glucosyltransferase(gtfs and gtfI) genes.";

RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL; AB026123; BAA77236.1; "

DR HSSP; P06278; IVS.

DR Interpro: IPR002479; CW\_binding.

DR Interpro: IPR003318; Glyco\_hydro\_70.

DR Pfam; PF04473; CW\_binding\_1; 10.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Plasmid.

SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 65.5%; Score 78; DB 2; Length 1338;

Best Local Similarity 71.4%; Pred. No. 0.0017;

Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHSTILEAWSNDNDPYLHD 21

DB 437 AIDHSTILEAWSNDNDYKVD 457

RESULT 9

ID 09LCJ7 PRELIMINARY; PRT: 1016 AA.

AC 09LCJ7;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE DEXTRANSUCRASE.

GN DSR1.

OS Leuconostoc mesenteroides.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

OC Leuconostoc.

OX NCBI\_TaxID=1245;

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL B-512F;

RX MEDLINE-20169623; PubMed-10705445;  
 RA Funane K., Mizuno K., Takahara H., Kobayashi M.;  
 RT "Gene encoding a dextranucrase-like protein in Leuconostoc  
 RL mesenteroides NRRL B-512F.";  
 DR EMBL: AB020020; BAA90527.1;  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1016 AA; 110343 MW; 8896FDE13CCBA7 CRC64;

Query Match 63.9%; Score 76; DB 2; Length 1016;  
 Best Local Similarity 61.9%; Pred. No. 0.0025;  
 Matches 13; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHLSILEWSDNDPTYLHD 21  
 :||:|||||:|||||:|  
 548 ANQHLSTLEWSDNDPEYKVD 568

RESULT 10

ID 09L466 PRELIMINARY; PRT: 1477 AA.

AC 09L466; 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE (EC 2.4.1.5).  
 GN DSRK.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B-1355;  
 RA Arguello-Morales M.A., Remaud-Simeon M., Pizut S., Sarcabal P.,  
 RA Willemot R.M., Monsan P.;  
 RT "Sequence analysis of the gene encoding alternansucrase, a sucrose  
 RT glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";  
 RL Submitted (OCF-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ250172; CAB76565.1;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Transferase: Glycosyltransferase.  
 SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFC831 CRC64;

Query Match 62.2%; Score 74; DB 2; Length 1477;  
 Best Local Similarity 66.7%; Pred. No. 0.0079;  
 Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 ANDHLSILEWSDNDPTYLHD 21  
 |||||||:|||||:|  
 DB 532 ANQHLSTLEWSDNDPEYKVD 552

RESULT 11

ID 052224 PRELIMINARY; PRT: 1508 AA.

AC 052224; 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-  
 DE GLUCOSYLTRANSFERASE).  
 GN DSRB.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B-1299;  
 RA Monchois V., Remaud-Simeon M., Monsan P., Willemot R.M.;  
 RL FEMS Microbiol. Lett. 0:0-0(1998).  
 CC -I- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
 CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
 DR EMBL: AF030129; AAB95453.1;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Transferase: Glycosyltransferase.  
 SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 62.2%; Score 74; DB 2; Length 1508;  
 Best Local Similarity 66.7%; Pred. No. 0.0081;  
 Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 ANDHLSILEWSDNDPTYLHD 21  
 |||||||:|||||:|  
 DB 563 ANQHLSTLEWSDNDPEYKVD 583

RESULT 12

ID 09EZHS PRELIMINARY; PRT: 1508 AA.

AC 09EZHS; 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE DSRB742.  
 GN DSRB742.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B-742CB;  
 RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
 RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF294469; AAG38021.1;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Transferase: Glycosyltransferase.  
 SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F7AE4F3A CRC64;

Query Match 62.2%; Score 74; DB 2; Length 1508;  
 Best Local Similarity 66.7%; Pred. No. 0.0081;  
 Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 ANDHLSILEWSDNDPTYLHD 21  
 |||||||:|||||:|  
 DB 563 ANQHLSTLEWSDNDPEYKVD 583

RESULT 13

ID 09LCH3 PRELIMINARY; PRT: 1575 AA.

AC 09LCH3; 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE.  
 GN GTPF.  
 OS Streptococcus oralis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.

OX NCBI\_TaxID=1303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC10537;  
 RX MEDLINE=20231779; PubMed=10768934;  
 RA Fujiwara T., Hoshino T., Oshima T., Sobue S., Hamada S.;  
 RT "Purification, characterization, and molecular analysis of the gene  
 encoding glucosyltransferase from *Streptococcus oralis*."  
 RL Infect. Immun. 68:2475-2483(2000).  
 DR EMBL; AB025228; BAA95201.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 17.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1575 AA; 176792 MW; 772A26E4D/C2E543 CRC64;

Query Match 61.3%; Score 73; DB 2; Length 1575;  
 Best Local Similarity 77.8%; Pred. No. 0.012;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 4 HLSILEAWSNDNDPYLHD 21  
 |||||  
 Db 549 HLSILEAWSNDNDPYNKD 566

RESULT 14  
 OS4178 PRELIMINARY; PRT; 1577 AA.  
 ID OS4178; OS4247;  
 AC OS4178; OS4247;  
 DT 01-NOV-1996 (TReMBLrel. 01. Created)  
 DT 01-NOV-1996 (TReMBLrel. 01. Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17. Last annotation update)  
 DE GLUCOSYLTRANSFERASE.  
 GN GTFG.  
 OS Streptococcus gordonii challis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 CC Streptococcus  
 OX NCBI\_TaxID=29390;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CHALLIS;  
 RX MEDLINE=96157084; PubMed=8586195;  
 RA Vickerman M.M., Sulavik M.C., Clewell D.B.;  
 RT "Molecular analysis of Streptococcus gordonii glucosyltransferase  
 phase variants."  
 RL dev. Biol. Stand. 85:309-314(1995).  
 RN [2]  
 SEQUENCE OF 1-96 FROM N.A.  
 STRAIN=CHALLIS;  
 RX MEDLINE=92276337; PubMed=1534326;  
 RA Sulavik M.C., Tardif G., Clewell D.B.;  
 RT "Identification of a gene, rgg, which regulates expression of  
 glucosyltransferase and influences the Spp phenotype of Streptococcus  
 gordonii Challis."  
 RL J. Bacteriol. 174:3577-3586(1992).  
 DR EMBL; U12643; AAC43483.1; -;  
 DR EMBL; M89776; AAA26969.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 18.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1577 AA; 177805 MW; 5AE032BDC5E08D18 CRC64;

Query Match 61.3%; Score 73; DB 2; Length 1577;  
 Best Local Similarity 77.8%; Pred. No. 0.012;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 4 HLSILEAWSNDNDPYLHD 21  
 |||||

Db 551 HLSILEAWSNDNDPYNKD 568

RESULT 15  
 OS42AR4 PRELIMINARY; PRT; 1527 AA.  
 ID OS42AR4  
 AC OS42AR4;  
 DT 01-MAY-1999 (TReMBLrel. 10. Created)  
 DT 01-MAY-1999 (TReMBLrel. 10. Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17. Last annotation update)  
 DE DEXTRANSUCRASE.  
 GN DEX.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 CC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-512-F;  
 RA Bhatnagar R., Singh D.K.S.;  
 RT "Cloning and Molecular Characterization of Dextranucrase Gene from  
 Leuconostoc mesenteroides NRRL B-512F."  
 RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; U81374; AAD10952.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 16.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW DEX.  
 SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 59.7%; Score 71; DB 2; Length 1527;  
 Best Local Similarity 66.7%; Pred. No. 0.024;  
 Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 1 ANDHLSILEAWSNDNDPYLHD 21  
 |||||  
 Db 581 ANQHLSILEAWSNDNDPYLHD 601

Search completed: March 27, 2002, 14:26:01  
 Job time: 1673 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: March 27, 2002, 13:57:52 ; Search time 198.55 Seconds  
(without alignments)  
7.088 Million cell updates/sec

Title: US-09-290-049a-2  
Perfect score: 95  
Sequence: 1 VPSYSTRAHDSVODLIA 19

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseqp/AA1980.DAT:\*  
2: /SID88/gcgdata/geneseq/geneseqp/AA1981.DAT:\*  
3: /SID88/gcgdata/geneseq/geneseqp/AA1982.DAT:\*  
4: /SID88/gcgdata/geneseq/geneseqp/AA1983.DAT:\*  
5: /SID88/gcgdata/geneseq/geneseqp/AA1984.DAT:\*  
6: /SID88/gcgdata/geneseq/geneseqp/AA1985.DAT:\*  
7: /SID88/gcgdata/geneseq/geneseqp/AA1986.DAT:\*  
8: /SID88/gcgdata/geneseq/geneseqp/AA1987.DAT:\*  
9: /SID88/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
10: /SID88/gcgdata/geneseq/geneseqp/AA1989.DAT:\*  
11: /SID88/gcgdata/geneseq/geneseqp/AA1990.DAT:\*  
12: /SID88/gcgdata/geneseq/geneseqp/AA1991.DAT:\*  
13: /SID88/gcgdata/geneseq/geneseqp/AA1992.DAT:\*  
14: /SID88/gcgdata/geneseq/geneseqp/AA1993.DAT:\*  
15: /SID88/gcgdata/geneseq/geneseqp/AA1994.DAT:\*  
16: /SID88/gcgdata/geneseq/geneseqp/AA1995.DAT:\*  
17: /SID88/gcgdata/geneseq/geneseqp/AA1996.DAT:\*  
18: /SID88/gcgdata/geneseq/geneseqp/AA1997.DAT:\*  
19: /SID88/gcgdata/geneseq/geneseqp/AA1998.DAT:\*  
20: /SID88/gcgdata/geneseq/geneseqp/AA1999.DAT:\*  
21: /SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:\*  
22: /SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	84	88.4	1592	14	AAAR32925
2	66	69.5	2057	21	AAAB10667
3	61	64.2	1577	17	AAAR91047
4	42	44.2	616	21	AAAB23861
5	42	44.2	2625	19	AAAM55887
6	42	44.2	2627	19	AAAM61347
7	41.5	43.7	713	17	AAAR9797
8	41	43.2	170	22	AAE06785
9	41	43.2	344	14	AAAR40918
10	41	43.2	552	22	AAE06772
11	41	43.2	552	22	AAE07043

12	41	43.2	662	22	AAAG2736	C glutamicum prote
13	41	43.2	1049	22	AAE04833	Human SCP006 phosph
14	41	43.2	1120	20	AAAY19780	B. burgdorferi ant
15	41	43.2	1123	21	AAAY58277	Helicobacterium
16	41	43.2	1146	20	AAAY19779	B. burgdorferi ant
17	40	42.1	149	21	AAAB40856	Human ORFX ORF620
18	40	42.1	170	22	AAE06786	Human dual-specific
19	40	42.1	241	22	AAE06776	Human dual-specific
20	40	42.1	359	22	AAAY06263	RetA protein of CT
21	40	42.1	431	11	AAAR07464	Polypeptide with e
22	40	42.1	431	20	AAAR87890	Protein encoded by
23	40	42.1	484	22	AAAB85363	Human phosphatase
24	40	42.1	498	22	AAE04835	Human SCP001 phosph
25	40	42.1	509	22	AAE06775	Human dual-specific
26	40	42.1	509	22	AAE07045	Human dual-specific
27	40	42.1	650	22	AAAB38706	Human polypeptide
28	40	42.1	759	22	AAAM40492	Human polypeptide
29	40	42.1	759	22	AAAM40493	Human polypeptide
30	39	41.1	174	22	AAE01115	Human gene 2 encod
31	39	41.1	356	22	AAAG82330	S. epidermidis ope
32	39	41.1	446	20	AAAY25684	Cockroach allergen
33	39	41.1	489	22	AAAG90051	C glutamicum prote
34	39	41.1	946	22	AAAG70686	S. cerevisiae apopt
35	39	41.1	2058	22	AAAB97070	Human polypeptide
36	38.5	40.5	237	16	AAAR84562	Human prostacyclin
37	38.5	40.5	310	22	AAAB94596	Human protein sequ
38	38.5	40.5	345	20	AAAY17246	Human protein kina
39	38.5	40.5	350	21	AAAB18877	A prolactin regula
40	38.5	40.5	417	21	AAAB18886	Human prolactin
41	38.5	40.5	417	21	AAAB94409	Human protein sequ
42	38.5	40.5	500	16	AAAR84561	Human prostacyclin
43	38.5	40.5	557	19	AAAY20666	Human neurofilamen
44	38	40.0	166	22	AAAM35026	Peptide #10063 enc
45	38	40.0	229	21	AAAG23447	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1	
ID	AAAR32925 standard; Protein; 1592 AA.
XX	
AC	AAAR32925;
XX	
DT	28-JUN-1993 (first entry)
XX	
DE	Glucosyltransferase I.
XX	
KW	GT-I; Streptococcus; dental; carries.
XX	
OS	Streptococcus sobrinus.
XX	
PN	JP05023188-A.
XX	
PD	02-FEB-1993.
XX	
PF	25-JUL-1991; 91JP-0186592.
XX	
PR	25-JUL-1991; 91JP-0186592.
XX	
PA	(FUKU/) FUKU I.
PA	(KATO/) KATO K.
XX	
DR	WPI; 1993-079449/10.
XX	
PT	N-PSDB; AAQ37760.
PT	DNA sequence glucosyl:transferase-I - comprises Streptococcus
PT	sobrinus DNA sequence with at least one nucleotide added or
XX	deleted
PS	Claim 13; Page 15; 29pp; Japanese.

CC The DNA sequence from *Streptococcus sobrinus* strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating  
 CC *S. sobrinus* 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with Sau3AI and fractionating on agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and *E. coli* JM109  
 CC transformed with it. A GT-1 expressing clone was isolated and  
 CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 CC  
 XX  
 SQ Sequence 1592 AA:

Query Match 88.4%; Score 84; DB 14; Length 1592;  
 Best Local Similarity 88.9%; Pred. No. 1.7e-06;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSTRAHDSVODLI 18  
 ||||| |||||  
 548 vpsysfaraahdsvegdll 565

## RESULT 2

AAB10667  
 ID AAB10667 standard; Protein; 2057 AA.

AC AAB10667;

DT 19-JAN-2001 (first entry)

DE L. mesenteroides alternan sucrose protein.

KW Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 KM syrup.

OS Leuconostoc mesenteroides.

PN DE19905069-A1.

PD 10-AUG-2000.

PF 08-FEB-1999; 99DE-1005069.

PR 08-FEB-1999; 99DE-1005069.

PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

Kossmann J, Welsh T, Quanz M, Knuth K;

DR WPI: 2000-550294/51.

DR N-PSDB: AAA97904.

PT New nucleic acid encoding recombinant *Leuconostoc mesenteroides*  
 PT alternan sucrose protein and methods of alternan and fructose  
 PT production

PS Claim 1a; Page 30-36; 64pp; German.

CC This invention describes a novel nucleic acid molecule (I) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group). The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The alternan and/or  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC starch. This sequence represents the *Leuconostoc mesenteroides* alternan  
 CC sucrose protein which is described in the method of the invention.

SQ Sequence 2057 AA;

Query Match 69.5%; Score 66; DB 21; Length 2057;  
 Best Local Similarity 61.1%; Pred. No. 0.0049;  
 Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSTRAHDSVODLI 18  
 :|:|:|:|:|:|:|:|:|:|  
 Db 757 ipnysfaraahdsvegdpl 774

## RESULT 3

AAR91047  
 ID AAR91047 standard; Protein; 1577 AA.

AC AAR91047;

DT 22-MAY-1996 (first entry)

DE Alpha-D-glucosyltransferase.

KW Alpha-D-glucosyltransferase; primer-independent; soluble glucan;  
 KM sucrose; transgenic plant; cloning; *Escherichia coli*;

KW phage lambda-cl3; vector; plasmid pGSG501; plasmid pGSG502;

KW gene transfer; crop improvement; storage carbohydrate; pasture;  
 KM feedstuff; senescence; dextran; binder; food; pharmaceutical.

OS *Streptococcus salivarius* strain ATCC 25975.

PN W09606173-A1.

PD 29-FEB-1996.

PF 24-AUG-1995; 95WO-AU00527.

PR 24-AUG-1994; 94AU-0007643.

PA (GIFE/) GIFFARD P M.

PA (JACO/) JACOUES N A.

PA (SIMP/) SIMPSON C L.

PI Giffard PM, Jacques NA, Simpson CL;

DR WPI: 1996-151376/15.

DR N-PSDB: AAT13139.

PT plants contg. new bacterial DNA encoding glucosyl transferase  
 PT activity - retain higher levels of stored carbohydrate(s) in a form  
 PT readily digestible by ruminants

PS Claim 4; Page 16-20; 31pp; English.

CC The sequence represents an alpha-D-glucosyltransferase from  
 CC *Streptococcus salivarius*. The enzyme is primer-independent, and  
 CC produces soluble glucan from sucrose. A gene encoding the enzyme  
 CC may be cloned and expressed in *Escherichia coli* using a subclone  
 CC of phage lambda-cl3, e.g. plasmid pGSG501 or plasmid pGSG502. The  
 CC DNA may also be expressed in a transgenic plant, to improve the  
 CC level of stored carbohydrate in a pasture plant which normally  
 CC contains low levels, or to prevent degradation of stored carbohydrate  
 CC during plant senescence. Dextran may be isolated from the plant, for  
 CC use as a food binder or pharmaceutical additive. Primer independence  
 CC ensures that the enzyme will be functional in plants. The glucan is  
 CC poorly degraded in plants but easily degraded by bacteria in the rumen  
 CC of grazing livestock.

SQ Sequence 1577 AA;

Query Match 64.2%; Score 61; DB 17; Length 1577;  
 Best Local Similarity 64.7%; Pred. No. 0.03;  
 Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 SYSFIRAHDSVQDLIA 19  
:|:|||||||::|  
Db 661 nyifvrahdsqvavla 677

# RESULT 4

AAB23861  
ID AAB23861 standard; Protein; 616 AA.

AC AAB23861;

DT 17-JAN-2001 (first entry)

DE Haemophilus influenzae adhesin (Hia) protein from NTHI strain 12.

KW Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine;  
KW non-typable Haemophilus influenzae; antiinflammatory; auditory;  
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;  
KW diagnosis; immunogenic; antigen.

Haemophilus influenzae.

PN WO200055191-A2.

PD 21-SEP-2000.

PF 16-MAR-2000; 2000WO-CA00289.

PR 16-MAR-1999; 99US-0268347.

PA (CONN-) CONNAUGHT LAB LTD.

PI Loosmore SM, Yang Y, Klein MH;

XX WPI; 2000-618897/59.

DR N-PSDB; AAA92500.

XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for  
PT use as antigens and vaccines and for treating Hemophilus influenzae  
PT infection.

PS Claim 1; Fig 25; 275pp; English.

CC The present sequence represents a Haemophilus influenzae adhesin (Hia)  
CC protein from the non-typable Haemophilus influenzae (NTHI) strain 12.  
CC Hia genes and proteins have antiinflammatory, auditory and antibacterial  
CC activities, and can be used in the production of a vaccine. An  
CC immunogenic composition comprising an Hia gene, a polypeptide encoded  
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing  
CC protection against disease caused by Haemophilus strains in a  
CC susceptible host, preferably a human. An Hia protein is useful as an  
CC antigen, in immunogenic preparations including vaccines, as a carrier  
CC for other immunogens, and in the generation of diagnostic reagents. Hia  
CC is useful for treating diseases caused by the infection of Haemophilus  
CC influenzae such as meningitis, epiglottitis, septicaemia and otitis  
CC media. Recombinant production of Hia favours high recovery of the  
CC protein compared to the low recovery of native protein from Haemophilus  
CC influenzae species. A truncated protein has a significantly higher  
CC amount of recovery than a full-length protein.

XX Sequence 616 AA;

Query Match 44.2%; Score 42; DB 21; Length 616;  
Best Local Similarity 38.9%; Pred. No. 32;

Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 PSYSFIRAHDSVQDLIA 19  
|:|:|:|:|:|:|:|  
Db 42 pdyatvraemgkvqelka 59

# RESULT 5

AAW55887  
ID AAW55887 standard; Protein; 2625 AA.

XX AAW55887;

DT 22-JUL-1998 (first entry)

DE Human telomerase.

KW Rat; telomerase; human; cancer; screening; inhibitor; elucidation;  
KW detection; probe; diagnosis; cell growth; ageing.

OS Homo sapiens.

PN WO9807838-A1.

PD 26-FEB-1998.

PF 21-AUG-1997; 97WO-JP02904.

PR 17-FEB-1997; 97JP-0031807.

PR 21-AUG-1996; 96JP-0219761.

PR 31-JAN-1997; 97JP-0018878.

PA (MITU) MITSUBISHI CHEM CORP.

PI Fujino Y, Harada N, Ishikawa F, Nakamura H, Takahashi K;

XX WPI; 1998-169149/15.

DR N-PSDB; AAV25989.

XX Claim 8; Page 74-93; 106pp; Japanese.

CC The present sequence represents a human protein component of telomerase.  
CC The DNA or RNA encoding the telomerase protein component or its  
CC fragments can be used as a nucleotide probe for the detection of  
CC cancer cells and for diagnosis of cancer. Potential telomerase  
CC inhibitors can be screened by measuring their effect on the assay  
CC of the active form in cells or tissues. The polypeptide and DNA  
CC coding for it can be used in the elucidation of biological control  
CC mechanisms of, e.g. cell growth or ageing and of the mechanisms of  
CC cancer development.

XX Sequence 2625 AA;

Query Match 44.2%; Score 42; DB 19; Length 2625;  
Best Local Similarity 47.1%; Pred. No. 1.8e+02;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVQDL 17  
:|:|:|:|:|:|:|

Db 205 mpysyslgeeevedl 221

# RESULT 6

AAW61347  
ID AAW61347 standard; Protein; 2627 AA.

XX AAW61347;

DT 12-OCT-1998 (first entry)

DE Human telomerase RNA interactive protein-1 (TRIP1).

XX TRIP1; human; telomerase RNA interactive protein-1; cancer; AIDS;  
XX ageing; therapy.

OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 1155 /note= "residue 1155 is given as Xaa in Fig 3
FT	(translates as Gln)"
FF	
PN	MO9812143-A1.
XX	
PD	
XX	22-MAY-1998.
PF	13-NOV-1997; 97WO-US21248.
XX	
PR	16-OCT-1997; 97US-0951733.
PR	15-NOV-1996; 96US-0871189.
XX	11-JUN-1997; 97US-0873039.
XX	
	(AMGE-) AMGEN CANADA INC.
	(AMGE-) AMGEN INC.
PI	Harrington LA, Robinson MO;
XX	
DR	WPI; 1998-297946/26.
DR	N-PSSB; AAV27865.
XX	
PT	New nucleic acid encoding human telomerase protein-2 - used for regulating telomerase activity, e.g. for treating cancer or acquired immune deficiency syndrome
PT	
XX	
PS	Example 2; Fig 3; 150pp; English.
XX	
CC	This polypeptide comprises human telomerase RNA interactive protein-1 (TRIP1). Its amino acid sequence was deduced from cDNA clones (see AAV27865) obtained from a colon tumour cell line LIM1863 cDNA library. The invention relates to novel genes encoding components of the telomerase enzyme complex, including TRIP1 and telomerase protein 2 (TP2, see also AAW61349). These polypeptides may be useful as therapeutic agents in those cases where increasing TRIP1 activity or TP2 activity is desired, e.g. for treatment of HIV infection, AIDS and ageing disorders. In situations in which TRIP1 and/or TP2 activity is to be decreased, such as in cancer cells in which TRIP1 activity and/or TP2 activity is elevated, TRIP1 and/or TP2 may serve as a target to identify a molecule which inhibits activity, or which decreases or inhibits the protein-protein interaction of TRIP1 and TP2, or the binding of either TRIP1 or TP2 to telomerase RNA. Alternatively, ex vivo or in vivo gene therapy may be used to administer TRIP1 or TP2 antisense molecules, or DNA constructs may serve to disrupt or enhance TRIP1 and/or TP2 expression in cells, and to create dominant negative inhibitors of TRIP1 or TP2.
CC	
CC	
SO	Sequence 2627 AA;
Query Match	44.2%; Score 42; DB 19; Length 2627;
Best Local Similarity	47.1%; Pred. No. 1.8e+02;
Matches 8; Conservative	3; Mismatches 6; Indels 0; Gaps 0.
OY	1 VPSFIRAHDESEVODL 17 :     :
Dd	205 mpysylsgieeedl 221
RESULT 7	
AAR99797	
ID	AAR99797 standard; Protein: 713 AA.
XX	
AC	AAR99797;
XX	
DT	26-NOV-1996 (first entry)
XX	
DE	Lysine decarboxylase (W3110) of E.coli.
XX	

KW	Lysine decarboxylase; L-lysine; cadA; transformant; deletion;
KW	mutant.
OS	Escherichia coli.
PN	WO9617930-A1.
PD	13-JUN-1996.
XX	
XX	05-DEC-1995; 95WO-JP02481.
PF	
XX	09-DEC-1994; 94JP-0306386.
PR	
PA	(AJIN ) AJINOMOTO CO INC.
XX	
PL	Kikuchi Y, Kojima H, Suzuki T;
XX	
DR	WPI; 1996-287175/29.
DR	N-PSDB; AAT34583.
XX	
PT	Production of L-lysine by culture of transformant Escherichia - in
PT	which expression of new lysine decarboxylase gene and or Cda gene
PT	has been partly or wholly suppressed
XX	
PS	Claim 1; Page 24-27; 45pp; Japanese.
CC	
CC	Transformant forms of Escherichia species (e.g. E.coli), in which
CC	the expression of the w1110 lysine decarboxylase gene (AAT34583), and/
CC	or the cda gene (AAT34584), have been partly or wholly suppressed by
CC	deletion of part or all of the gene may be used for the efficient
CC	production of L-lysine when cultured.
XX	
SQ	Sequence 713 AA;
XX	
Query Match	43.7%; Score 41.5; DB 17; Length 713;
Best Local Similarity	45.0%; Pred. No. 47;
Matches 9; Conservative 3; Mismatches 5; Indels 3; Gaps 1.	
QY	1 VPSYSFIRAH--DSEVDL 17
	: : :
Db	77 lplyaflnthslmdsvqdm 96
XX	
RESULT 8	
AAE06785	
ID	AAE06785 standard; Protein; 170 AA.
XX	
AC	AAE06785;
XX	
DT	16-OCT-2001 (first entry)
DE	
XX	
XX	Human dual-specificity phosphatase (DSP)-12 partial protein.
XX	
KW	Human; dual-specificity phosphatase, DSP-12; dual phosphorylation motif;
KW	mitogen-activated protein kinase; MAP-kinase; graft-versus-host disease;
KW	GvHD; Duchenne muscular dystrophy; therapy; cancer; autoimmune disease;
KW	allergy; metabolic disease; cell growth; cell proliferation; cytostatic;
KW	cell cycle abnormality; cell differentiation; antiallergic; muscular;
KW	immunosuppressive.
XX	
OS	Homo sapiens.
XX	
PN	WO200157221-A2.
XX	
PD	09-AUG-2001.
XX	
PF	01-FEB-2001; 2001WO-US03429.
XX	
PR	02-FEB-2000; 2000US-0179886.
XX	
PA	(CEPT-) CEPTYR INC.
XX	



P1	Luche RM, Wei B;
XX	
DR	WPI: 2001-486887/53.
XX	
PT	New isolated dual-specificity phosphatase polypeptide for treating
PT	cancer, graft-versus-host disease, autoimmune diseases, allergies,
PT	metabolic diseases, abnormal cell growth and abnormal cell
PT	proliferation -
XX	
PS	Example 1; Fig 6; 81pp; English.
XX	
CC	The present sequence is human dual-specificity phosphatase (DSP)-12
CC	patrial protein. Inactivation of mitogen-activated protein kinase (MAP-
CC	kinase) is mediated by dephosphorylation at a dual phosphorylation motif
CC	by DSP which is referred to as MAP-kinase phosphatase. An agent that
CC	modulates DSP is useful for treating a disorder selected from Duchenne
CC	muscular dystrophy, cancer, graft-versus-host disease (GVHD), autoimmune
CC	diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
CC	cell proliferation and cell cycle abnormalities. DSP is useful for
CC	identifying antibodies and other agents that inhibit DSP-12 and/or
CC	DSP-13 activity. DSP and the agents identified are useful for
CC	modulating cell proliferation, differentiation and survival. DSP is
CC	useful in screening assays for modulators of enzyme activity and
CC	substrate binding and for dephosphorylating a substrate of DSP-12 or
CC	DSP-13.
SQ	Sequence 170 AA;
XX	
XX	
Query Match	43.2%; Score 41; DB 22; Length 170;
Best Local Similarity	47.1%; Pred. NO. 11;
Matches	8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
OY	3 SYSFIRAHDSFYODLIA 19
	:     :         :
Db	72 aynhlrvydeetdlla 88
RESULT	9
AAR40918	
ID	AAR40918 standard; Protein; 344 AA.
XX	
AC	AAR40918;
DT	
XX	21-FEB-1994 (first entry)
DE	Bacillus subtilis inositol dehydrogenase.
XX	
KW	Recombinant protein production; inositol dehydrogenase. Bacillus subtilis.
PM	JP05192163-A.
XX	
PD	03-AUG-1993.
XX	
PF	30-SEP-1991; 91JP-0252073.
XX	
PR	30-SEP-1991; 91JP-0252073.
PA	(SUNR ) SUNTORY LTD.
XX	
DR	WPI: 1993-277476/35.
DR	N-PSDB; AAO48539.
XX	
PT	Inositol dehydrogenase gene - isolated from e.g. bacteria
PT	belonging to Bacillus subtilis, in large amt.
XX	
PS	Claim 2; Page 2; 18pp; Japanese.
XX	
CC	The Bacillus subtilis inositol dehydrogenase gene has been cloned
CC	and sequenced. Microorganisms transformed by the coding sequence can
CC	be cultured to produce large amounts of recombinant inositol

CC	dehydrogenase.
XX	
SQ	Sequence 344 AA;
QY	6 FIRAHDSVQDLI 18   :     :       276 fvaaydveldgfdl 288
Db	
RESULT 10	
ID AAE06772	AAE06772 standard; Protein: 552 AA.
XX	
AC AAE06772:	
XX	
DT 16-OCT-2001	(first entry)
XX	
DE Human dual-specificity phosphatase (DSP)-12 protein.	
KW Human: dual-specificity phosphatase; DSP-12; dual phosphorylation motif; KW mitogen-activated protein kinase; MAP-kinase; graft-versus-host disease; KW GvHD; Duchenne muscular dystrophy; therapy; cancer; autoimmune disease; KW allergy; metabolic disease; cell growth; cell proliferation; cytostatic; KW cell cycle abnormality; cell differentiation; antiallergic; muscular; KW immunosuppressive.	
KX Homo sapiens.	
OS	
XX	
Key Location/Qualifiers	
FH Domain 249..269	/note= "Active site domain"
FT Misc-difference 386	/note= "Amino acid Thr is present at this location in the sequence shown in sequence listing of the specification"
FT FT	
XX	
PN WO200157221-A2.	
XX	
PD 09-AUG-2001.	
XX	
PF 01-FEB-2001; 2001WO-US03429.	
XX	
PR 02-FEB-2000; 2000US-0179886.	
XX	
PA (CEPT-) CEPTR INC.	
XX	
Luche RM, Wei B;	
PI WPI: 2001-488887/53.	
DR N-PSDB; AAD12965.	
XX	
PT New isolated dual-specificity phosphatase polypeptide for treating PT cancer, graft-versus-host disease, autoimmune diseases, allergies, PT metabolic diseases, abnormal cell growth and abnormal cell PT proliferation -	
XX	
Claim 1; Fig 2; 81pp; English.	
XX	
CC The present sequence is human dual-specificity phosphatase (DSP)-12 CC protein. Inactivation of mitogen-activated protein kinase (MAP-kinase) CC is mediated by dephosphorylation at a dual phosphorylation motif by DSP CC which is referred to as MAP-kinase phosphatase. An agent that modulates CC DSP is useful for treating a disorder selected from Duchenne muscular CC dystrophy, cancer, graft-versus-host disease (GvHD), autoimmune diseases, CC allergies, metabolic diseases, abnormal cell growth, abnormal cell CC proliferation and cell cycle abnormalities. DSP is useful for CC identifying antibodies and other agents that inhibit DSP-12 and/or CC DSP-13 activity. DSP and the agents identified are useful for	

CC modulating cell proliferation, differentiation and survival. DSP is  
CC useful in screening assays for modulators of enzyme activity and  
CC substrate binding and for dephosphorylating a substrate of DSP-12 or  
CC DSP-13.  
XX  
SQ Sequence 552 AA;

Query Match 43.2%; Score 41; DB 22; Length 552;  
Best Local Similarity 47.1%; Pred. No. 43;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 3 SYSFIRAHDEYVDLIA 19  
: | | | | | | | | | |  
Db 214 aynhlrvydeetdlla 230

## RESULT 11

AAE07043 standard; Protein: 552 AA.

AAE07043;

16-OCT-2001 (first entry)

Human dual-specificity phosphatase (DSP)-12 mutant protein, C253S.

Human: dual-specificity phosphatase; DSP-12; dual phosphorylation motif;  
mitogen-activated protein kinase; MAP-kinase; graft-versus-host disease;  
GVHD; Duchenne muscular dystrophy; therapy; cancer; autoimmune disease;  
allergy; metabolic disease; cell growth; cell proliferation; cytostatic;  
cell cycle abnormality; cell differentiation; antiallergic; muscular;  
immunosuppressive; mutant; mutein; variant.

Homo sapiens.  
Synthetic.

Key Location/Qualifiers

Misc-difference 253 /note= "Wild type Cys substituted with Ser"

WO200157221-A2.

09-AUG-2001.

01-FEB-2001; 2001WO-US03429.

02-FEB-2000; 2000US-0179886.

(CEPT ) CEPTYR INC.

Luche RM, Wei B;

WPI; 2001-488887/53.

New isolated dual-specificity phosphatase polypeptide for treating  
PT cancer, graft-versus-host disease, autoimmune diseases, allergies,  
PT metabolic diseases, abnormal cell growth and abnormal cell  
PT proliferation

Disclosure; Page -: 81pp; English.

The present sequence is human dual-specificity phosphatase (DSP)-12  
mutant protein, C253S. Inactivation of mitogen-activated protein kinase  
(MAP-kinase) is mediated by dephosphorylation at a dual phosphorylation  
motif by DSP which is referred to as MAP-kinase phosphatase. An agent  
that modulates DSP is useful for treating a disorder selected from  
CC Duchenne muscular dystrophy, cancer, graft-versus-host disease (GVHD),  
CC autoimmune diseases, allergies, metabolic diseases, abnormal cell growth,  
CC abnormal cell proliferation and cell cycle abnormalities. DSP is useful  
CC for identifying antibodies and other agents that inhibit DSP-12 and/or  
CC DSP-13 activity. DSP and the agents identified are useful for  
modulating cell proliferation, differentiation and survival. DSP is

CC useful in screening assays for modulators of enzyme activity and  
CC substrate binding and for dephosphorylating a substrate of DSP-12 or  
CC DSP-13.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the DSP-12 sequence (AAE06772) given as SEQ ID NO: 2 in  
CC figure 2.  
XX  
SQ Sequence 552 AA;

Query Match 43.2%; Score 41; DB 22; Length 552;  
Best Local Similarity 47.1%; Pred. No. 43;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 3 SYSFIRAHDEYVDLIA 19  
: | | | | | | | | | |  
Db 214 aynhlrvydeetdlla 230

## RESULT 12

AAE07043 standard; Protein: 662 AA.

AAE07043;

26-SEP-2001 (first entry)

C glutamicum protein fragment SEQ ID NO: 6490.

Corynebacterium; amino acid synthesis; vitamin; saccharide;  
organic acid synthesis.

Corynebacterium glutamicum.

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOW ) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI; 2001-376931/40.

N-PSDB; AAH67955.

Novel polynucleotides derived from Corynebacterium bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

SQ Sequence 662 AA;

Query Match 43.2%; Score 41; DB 22; Length 662;  
 Best Local Similarity 53.3%; Pred. NO. 53;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 5 SFIRAHDESEVODLIA 19  
 | : | | | | | : |  
 Db 415 siekhdspaqdlva 429

RESULT 13  
 ID AAE04833 standard; Protein: 1049 AA.  
 AC AAE04833;  
 DT 10-SEP-2001 (first entry)  
 XX

Human SGP006 phosphatase polypeptide.

Human; SGP006 phosphatase polypeptide; phosphatase-related disease;  
 immune-related disorder; ocular disease; organ transplant rejection;  
 infection; diabetes; pain; sexual dysfunction; Alzheimer's disease;  
 metabolic disorder; haematopoietic cancer; mood disorder; cardiac;  
 Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 cardiovascular disease; brain; neuronal-associated disease; dyskinesia;  
 attention disorder; cognitive disorder; psychotic disorder; cytostatic;  
 neurological disorder; vincristine; neurotropic; cerebroprotective; therapy;  
 neuroprotective; antibacterial; vulvar; transglutinin; antiasthmatic;  
 hypotensive; immunosuppressive; antiproliferative; angiogenic; hypertensive;  
 antifungal; dual specificity phosphatase; DSP; MAP kinase phosphatase;  
 MKP; mgraine; chromosome 12q21.3-q22.

XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 1..173  
 FT /label= Catalytic\_domain  
 FT 308..446  
 FT /label= Phosphatase\_domain  
 XX  
 PN MO200146394-A2.  
 XX  
 PD 28-JUN-2001.  
 XX  
 PF 21-DEC-2000; 2000MO-US34736.  
 XX  
 PR 21-DEC-1999; 99US-0173255.  
 XX 28-DEC-1999; 99US-0175766.  
 XX 25-JAN-2000; 2000US-0178078.  
 PR 31-JAN-2000; 2000US-0179301.  
 XX  
 PA (SUGEN-) SUGEN INC.  
 XX  
 PI Plowman GD, Martinez R, Whyte D, Manning G, Sudarsanam S, Hill RJ,  
 PI Flanagan P;  
 XX  
 DR WPI: 2001-418058/44.  
 DR N-PSDB: AAD09491.  
 XX  
 PT Novel phosphatase polypeptide useful for treating cancers,  
 PT immune-related diseases and disorders, cardiovascular disease, brain or  
 PT neuronal-associated diseases and metabolic disorders -  
 XX  
 PS Claim 7; Fig 2; 186pp; English.  
 XX  
 CC The present invention relates to phosphatase polypeptides, nucleotide  
 CC sequences encoding them, as well as various products and methods useful  
 CC for the diagnosis and treatment of various phosphatase-related diseases  
 CC and conditions. Substance that modulates the activity of phosphatase  
 CC polypeptide is used to treat immune-related diseases and disorders,

CC cardiovascular disease, brain or neuronal-associated diseases and  
 CC metabolic disorders, including cancers of tissues, cancers of  
 CC haematopoietic origin, diseases of central and peripheral nervous  
 CC system, Alzheimer's disease, Parkinson's disease, multiple sclerosis,  
 CC amyotrophic lateral sclerosis, viral infections, infections caused by  
 CC prions, bacteria and fungi, ocular diseases, diabetes, migraines, pain,  
 CC sexual dysfunction, mood disorders, attention disorders, cognitive  
 CC disorders, hypotension, hypertension, psychotic disorders, neurological  
 CC disorders, dyskinesias and organ transplant rejection. The present  
 CC amino acid sequence is human SGP006 phosphatase polypeptide. This  
 CC sequence is classified as dual specificity phosphatase (DSP) and MAP  
 CC kinase phosphatase (MKP). SGP006 gene maps to chromosomal position  
 CC 12q21.3-q22.  
 XX  
 SQ Sequence 1049 AA;

Query Match 43.2%; Score 41; DB 22; Length 1049;  
 Best Local Similarity 47.1%; Pred. NO. 91;  
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 3 SYSFIRAHDESEVODLIA 19  
 : | | | | | : |  
 Db 354 ayhmrvydeetdlia 370

RESULT 14  
 AAY19780  
 ID AAY19780 standard; Protein: 1120 AA.  
 XX  
 AC AAY19780;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE B. burgdorferi antigenic protein, t12.aa.  
 XX  
 KM Antigenic protein; vaccine; Lyme disease; infection; detection.  
 XX  
 OS Borrelia burgdorferi.  
 XX  
 PN WO9859071-A1.  
 XX  
 PD 30-DEC-1998.  
 XX  
 PF 18-JUN-1998; 98WO-US12718.  
 XX  
 PR 03-SEP-1997; 97US-0057483.  
 XX 20-JUN-1997; 97US-0050359.  
 PR 22-JUL-1997; 97US-0053344.  
 PR 22-JUL-1997; 97US-0053377.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMUNE INC.  
 XX  
 PI Choi GH, Ewin AL, Hanson MS, Lathigra R;  
 XX  
 DR WPI: 1999-189980/16.  
 DR N-PSDB: AAX61476.  
 XX  
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the diagnosis, prevention and treatment of diseases  
 PT caused by Borrelia, particularly Lyme disease  
 XX  
 PS Claim 12; Page 57; 275pp; English.  
 XX  
 CC This sequence represents a Borrelia burgdorferi (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
 CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the Borrelia genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the Borrelia genus. The products can also  
 CC be used for detection of members of the Borrelia genus.

SQ Sequence 1120 AA;

Query Match 43.2%; Score 41; DB 20; Length 1120;  
 Best Local Similarity 46.7%; Pred. No. 99;  
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 YSFIRAHDSVODLI 18  
 |||: |||: |:  
 Db 740 ysfnkkydsdvkslv 754

RESULT 15

AAV58277

ID AAV58277 standard; Protein; 1123 AA.

XX AAV58277;

AC AAV58277;

27-MAR-2000 (first entry)

DE Heliothis armigera entomopoxvirus (HaEPV) DNA polymerase.

KM Entomopoxvirus; EPV; insect virus; intergenic region; heterologous DNA;  
 KW Insecticide; recombinant protein production; DNA polymerase; HaEPV.

OS Heliothis armigera entomopoxvirus.

XX WO9963062-A1.

XX 09-DEC-1999.

XX 28-MAY-1999; 99WO-AU00416.

XX 29-MAY-1998; 98AU-0003805.

PA (CSTR ) COMMONWEALTH SCI & IND RES ORG.

XX Dali DJ;

DR WPI: 2000-072880/06.

DR N-PSDB; AAV58277.

PT Infectious recombinant virus for use as insecticides or for expressing  
 therapeutic proteins

PS Examples; Fig 1b; 84pp; English.

This sequence represents Heliothis armigera entomopoxvirus (HaEPV)  
 DNA polymerase. The invention relates to infectious, spindle  
 body-producing entomopox viruses which have heterologous DNA inserted  
 in a 771 bp intergenic region between the fusolin and p68 genes.  
 CC Recombinant viruses containing heterologous DNA encoding  
 an insecticidal protein may be used for the control of insect pests.  
 CC Alternatively, viruses containing sequences encoding therapeutically  
 useful proteins such as interferon, human growth hormone or insulin may  
 be used for recombinant production of these proteins.

SQ Sequence 1123 AA;

Query Match 43.2%; Score 41; DB 21; Length 1123;

Best Local Similarity 58.3%; Pred. No. 99;

Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PSYSFIRAHDS 13

Db 684 pdysyllindke 695

Search completed: March 27, 2002, 13:57:53  
 Job time: 521 sec